International Journal of Research on Service-Learning & Community Engagement

Volume 12 | Issue 1

Role of Design for Causal Analysis in Research on Community Engagement

Review and Recommendations

Thomas A. Dahan
School of Nursing–Camden, Rutgers, The State University of New Jersey, United States

Recommended Citation
Dahan, T. A. (2024). Role of Design for Causal Analysis in Research on Community Engagement. Review and Recommendations. *International Journal of Research on Service-Learning and Community Engagement*, *12*(1). https://doi.org/10.37333/001c.127546

International Journal of Research on Service-Learning and Community Engagement
Volume 12, Issue 1 | 2024 | ISSN: 2374-9466 | https://ijrslce.scholasticahq.com/

Role of Design for Causal Analysis in Research on Community Engagement

Review and Recommendations

El papel del diseño para el análisis causal en la investigación sobre la participación comunitaria

Reseña y recomendaciones

Thomas A. Dahan

Abstract
The purpose of this review is to advance the application of causal inference strategies to service-learning and community engagement and offers recommendations for both practitioners and researchers. The review offers an introduction to various techniques for yielding causal conclusions and discusses an example of the technique from the literature. The conclusion offers recommendations for pursuing causal inference and improving research designs related to service-learning and community engagement.

Keywords: *causal inference, community engagement, research design, service-learning*

Thomas A. Dahan

Resumen
El propósito de esta revisión es ofrecer recomendaciones para profesionales e investigadores en cómo mejorar la aplicación de estrategias de inferencia causal en el aprendizaje y servicio y otras formas de participación comunitaria. Esta revisión ofrece una introducción de varias técnicas para llegar a conclusiones causales y además analiza un ejemplo de la técnica en la literatura existente. Se incluyen recomendaciones para buscar inferencias causales y mejorar los diseños de investigación relacionados con el aprendizaje y servicio y otras formas de participación comunitaria.

Palabras clave: *inferencia causal, compromiso comunitario, diseño de investigación, aprendizaje-servicio*

Editors' Note: Translation provided by ***Karla Diaz***
Instituto de Aprendizaje y Servicio
Universidad San Francisco de Quito
Ecuador

The credibility revolution in econometrics has generated strategies for research designs that have explored a variety of topics and strongly influenced policy decisions related to education and society (Angrist & Pischke, 2010). However, research on service-learning and community engagement (SLCE) less frequently applies methods that yield causal conclusions about the research whether focused on students, practitioners, or communities. Traditional causal designs rely heavily on experimental methods featuring randomization to assignment, which are rare in service-learning research (Steinberg et al., 2013). However, experimental methods are not the only designs that yield causal interpretations. This review focuses on contributions to research that employ a variety of causal inference strategies from the econometric tradition for research on service-learning and community engagement between 2005 and 2023.

For this review, I define SLCE as an educational approach that purposively engages diverse agents and institutions in collaborative partnership to address needs identified by and with community members (who are themselves agents in these relationships). This definition is inclusive of K-12, higher education, and non-formal practices of SLCE, but I acknowledge many of the examples drawn within are applications within higher education or are studies of higher education institutions and/or their SLCE work. Although the approaches of service-learning and community engagement are conceptually distinct, the research methods explored in this review can be applied to both.

This paper is organized as follows: I begin with a positionality statement to offer readers a reflexive account of my motivations and perspectives regarding this review. Then, I introduce the epistemological foundations of causal inference and discuss the role of two paragons of research design who deeply influence this topic. Next, I present a review of research that uses traditional experimental methods to achieve causal conclusions, introduce emerging econometric methods, and provide examples of their application to the field of research on community engagement. I conclude with recommendations to provide a path for future researchers who are designing research on community engagement. Finally, in addition to this review, I include an Appendix containing a glossary of terms used in this review and others germane to the topic to aid readers with the complicated language of econometrics and design for causal inference.

Causal Inference: Epistemology and Designs

Author's Positionality Statement

The positionality statement is a researcher's tool for identifying the assumptions and predispositions of the researcher in advance of engaging in the research process (Jamieson et al., 2023). The reflexivity of the positionality statement is at the core of the social constructivist, qualitative paradigm, but is increasingly incorporated into quantitative research. For myself, I wish to engage in reflective practice in regard to this work and open this section with a discussion of my own principles.

In the section following this statement, I only briefly discuss the problematic epistemic foundations of causal inference. I acknowledge that embedded in the belief in causality relationships is a predisposition to objectivity and reliability, divorcing the researcher from the contexts where community engagement work takes place and presupposes that students or communities want to be engaged in activities that involve manipulations to test for causal implications. Part of the motivation for this paper is to highlight potential methods that use observational data to draw conclusions that can be causal in nature without the baggage of experimentation often associated with the supposed "gold standard" of research design (Scriven, 2008).

In addition, I am motivated by the discussion of the need for research on and for equity in access, participation, and outcomes from SLCE. I advance the arguments of this paper with humility, acknowledging that not all practices in SLCE research can or should apply a causal inference framework. In particular, the critical work of advancing equity may favor other intersubjective and emancipatory methods that are difficult to reconcile with the quantitative approach to research methods. Discussion of equity outcomes and processes are often clouded in language regarding "heterogeneity of treatment effects" when explored and discussed with econometric methods.

I must also acknowledge that there are other causal inference frameworks I omitted from this review, such as Maxwell's (2012) approach to causal inference with qualitative research. Despite my own limitations in discussing qualitative approaches to causal inference, I also strive to make space for quantitative researchers in SLCE to advance the quality of research designs proposed and pursued when studying the practices of engagement. While I am interested in causal analysis, I also largely omit the path analysis tradition of structural equation modeling of causal inference from this review (Blalock, 2018). In particular, I want to introduce readers to a specific set of econometric techniques that may be applied to SLCE research.

Epistemological Foundations for Causal Inference

The epistemic foundation of causal inference is in David Hume's (orig. 1739, 2007) *A Treatise of Human Nature*, which is largely seen as an argument for empiricism, originally published in the 18th century. By the 19th and 20th centuries, these ideas served as the basis for a philosophy of science called positivism. The positivists believed that the only meaningful assertions were those that could be verified through scientific observations (Hempel, 1950).

By the middle of the 20th century, a number of philosophical movements set out to critique and reshape positivism. The qualitative/interpretivist paradigm of research emerged and critiqued the reductionism of the positivist movement (Denzin, 2008). Critical rationalism emerged as a reaction to and solution to the verificationism of positivist thinking (Popper, 1979). Critical theorists and post-modernist critiques of society and its structure also introduced new perspectives to and understanding of the social world, further complicating the philosophies of positivism (Mitchell & Rost-Banik, 2017).

The core philosophy of causal inference remained intact despite the epistemic criticisms of the positivist paradigm with which it is largely associated. Causal inference remains a philosophical proposition, regardless of the research methods applied to uncover the conclusions. For example, Cook and Campbell argue in their well-regarded methods text, that their approach is "derived from Mill's inductivist canons, a modified version of Popper's falsificationism, and a functionalist analysis of why cause is important in human affairs" (1979, as cited in Shadish, 2010). The approach of Campbell is intrinsically tied with the concepts of internal and external validity, with the internal validity of treatments being among the most important aspects of causal conclusions.

Other philosophies of causal inference take a counterfactual approach to causality. Much of the emergent work in causal inference builds from the counterfactual philosophy of causation, often associated with Donald Rubin (Imbens & Rubin, 2010). In this approach, the treatment condition and its counterfactual are estimated by observing outcomes from groups of individuals, with average differences between the treated and counterfactual interpreted as the average treatment effect on the treated. In this formulation, researchers accept limitations in their ability to argue an externally valid and generalizable treatment effect but argue the strength of their designs to develop internally valid conclusions that are observable in real-world settings increases the plausibility of the findings in settings beyond the research setting in which the finding is obtained.

Traditional Causal Inference Strategies Applied to Community Engagement

Three key assumptions must be tenable to interpret any research results to be causal: ignorable treatment assignment, exclusive potential outcomes, and monotonicity. These terms (and others) are each defined in the accompanying Appendix glossary of terms. Traditional causal designs rely heavily on experimental methods featuring randomization to assignment, which are rare in service-learning research (Steinberg et al., 2013). Randomization, when successful, enables results to be interpreted as causal because differences between treated and control groups are purely chance (which is the definition of ignorable treatment assignment). After the treatment is applied, the differences in the groups can be attributed to the treatment, as long as we can assume the treatment outcomes are exclusive to the group assigned to treatment and no one purposely defied their assignment to the treatment or control (i.e., monotonicity). This tradition of causal inference draws heavily from the Cook and Campbell tradition (Shadish, 2010).

The seminal work by Markus et al. (1993) was notable in the early research on service-learning because of its use of a strong quasi-experimental design including "blind assignment" to treatments. Osborne et al. (1998) replicated the approach. More recently, Brown (2011) compared social dominance outcomes among students randomly assigned to be engaged in service-learning and those who were assigned research on out-groups that are often the focus of service-learning activities. Pong and Lam (2023) used experimental random assignment to examine service-learning's effects on emotional intelligence quotient.

Randomization does not need to occur at the individual level for causal inference. Designs where the unit of analysis is at an aggregate unit such as classroom, university, or school district are commonly applied to overcome the potential for the treated and untreated units to interact with each other, which contaminates the treatment-control contrast (Murnane & Willett, 2010). This contrast is alternatively referred to as the "stable unit treatment value assumption," which is necessary for propensity score methods to be implemented (Guo & Fraser, 2014; Hernan & Robins, 2020). Because this assumption is so critical to successful randomized experiments, it is often a reason these techniques are difficult to implement in practice. But even in instances where randomized experiments have issues, the use of instrumental variables (discussed below) can recover the causal effects (Murnane & Willett, 2010).

While experimental design continues to be the gold standard to produce both internally and externally valid causal inference, it remains difficult to justify or scale randomized trials using community-engaged methods. First, the use of randomization in community settings may be impractical when the engagement is designed to be intentional, reciprocal, and lasting (Goodkind et al., 2017). Potential ethical concerns regarding the practices of "helicopter researchers" who enter communities with the interest of extracting research data and leaving the community without providing direct or indirect benefit from their knowledge creation are associated with the legacy of positivist research (Danley et al., 2022). Other existing critiques of quantitative methods applied to SLCE research are also worthy of acknowledgment regarding the general inadequacy of these methods to describe the complexity of phenomena of SLCE experiences (Jones & Foste, 2016; Shumer, 2000).

While this paper does not attempt to reconcile most issues related to quantitative methods in SLCE research, I will try to address the particular challenges related to randomized designs by introducing a set of alternative methods that yield potentially causal conclusions without randomization of assignment. Fortunately, there are other quasi-experimental methods that can yield defensible conclusions about causal effects that may be more effective or practical in applied community engagement settings than the randomized field experiment.

Commonly Applied Methods for Non-Experimental Data

Many researchers use multiple regression techniques to estimate the average treatment effect. For example, 29 articles in the *Michigan Journal of Community Service Learning* applied regression methods to estimate impacts. This technique is flexible (meaning it can be applied to a variety of outcome measures) and taught in many graduate programs.

Multiple regression techniques impose assumptions about how data are generated even when the assumptions are untenable by modeling the data as the association between the observed covariates in the model and the outcome, relegating the unexplained variation to its residual error. Using regression to identify causal effects embeds the assumptions by "holding constant" differences on covariates, enabling the association between a treatment indicator and the outcome to be interpreted as the average effect of the treatment (Morgan & Winship, 2015).

Two issues with this approach exist. First, suppose the values of the observed covariates in the total sample are different between the treated group and the untreated group. In that case, the regression technique ignores these systematic differences and estimates potentially extrapolated values that are not observed. Even worse, if an unobserved variable associated with the outcome and any covariate in the model is left out, it will create an association between the residual error and the treatment of interest, resulting in an omitted variable bias in an unknown direction. For example, Bowman et al. (2015) explore the longitudinal civic outcomes of college student involvement, offering a comparison of a multilevel regression model against a propensity score approach. The estimates for the regression techniques overestimate the effects by between 10 and 100 percent, seriously undermining the credibility of the regression-based estimator.

Many authors justifiably report this second potential omitted variable bias as simply a limitation of their research, acknowledging that it is virtually impossible to control all unobserved variables or relying on the argument that the regression demonstrates only association and not causation (Blalock, 2018). This conclusion is true but does not need to limit the ability to produce accurate and defensible conclusions about research results. It is this very problem that generated the credibility revolution in the first place (Angrist & Pischke, 2010; Leamer, 1983). The remainder of this paper advocates designs that enable the three key assumptions introduced above to be intentionally imposed by design, yielding results that improve upon multiple regression methods.

Emerging Causal Inference Strategies for Research on Community Engagement

Experimental methods are not the only designs that yield causal interpretations and multiple regression techniques can be improved upon because "you cannot fix by analysis what you bungled by design" (Light et al., 2009). Many emerging techniques for causal inference draw upon the epistemic foundations of the Rubin causal model applying counterfactual logic.

For example, natural experiments can offer "as good as random" assignment to treatment conditions. A variety of methods such as fixed effects longitudinal designs (Dahan, 2020) and difference in differences (Walcott et al., 2018) can be applied to natural experiments to yield defensible causal interpretations. Beyond the natural experiment, an increasingly popular method of estimating causal effects in service-learning research is through the propensity score analysis approach. Exemplars include Song et al. (2017); Maruyama et al. (2018); Soria et al. (2019); Schulzetenberg et al. (2020). Other methods that can generate causal conclusion include instrumental variables/fuzzy regression discontinuity design for the complier's average causal effect (Mo et al., 2022) and the synthetic control method (Pearl et al., 2013).

Fixed Effects and Difference in Differences

Conceptually, fixed effects and difference in differences make use of observations of individual units over time to estimate the causal relationship between a predictor and an outcome variable. The design regresses the time-demeaned outcome and predictor variables, which removes all time invariant heterogeneity (both observed and unobserved) that may take the form of a left-out variable bias. The remaining variation between the outcome and predictor can be understood to be a plausibly causal relationship between the two variables if we can assume the following: (1) the relationship between the predictor and the residual error is random (there are no other unobserved time-varying relationships confounding the predictor); (2) there is sufficient variation in the predictor relationship such that it is not collinear with itself across time (otherwise the observation would drop from the equation as part of the fixed effect).

An example of fixed effects in the SLCE literature is Dahan's (2020) study of the influence of the Learn and Serve America policy on community-level social capital outcomes. Social capital is an important mechanism to address problems of public concern and is theoretically linked to the practice of service-learning and community engagement. Dahan collected indicators of social capital over time for communities across the United States and sought to link the changes in these indicators to the growth of the Campus Compact and the subsequent reduction in its size after the federal government removed support for Learn and Serve America in 2011. By using the sharp discontinuity of the defunding as a natural experiment that indicates the changes between the period of growth and decline of the compact, he demonstrated that the policy had small but measurable effects equivalent to about half of a standard deviation of within-community changes in social capital (a moderate effect size for an outcome variable with fairly limited variation within places across time).

The treatment effect of interest in the fixed effects and difference-in-difference design is the average treatment effect on the treated (ATT). In contrast to true experimental designs, which estimate the average treatment effect (ATE) that is generalizable, the ATT is limited to the effect of the treatment on the units where the treatment status changes (the observations, where treatments are always present or no treatment is present, in any period drop from the equation as part of the fixed effect). While the outcome is based on the counterfactual outcomes of the units when they are untreated, we cannot claim that the ATT would have a valid causal impact for the always treated or untreated units, only for the units where treatment status changes.

With this delimitation, the design can be applied in situations where changes (such as curricular changes) are made somewhat abruptly, changing one group's treatment status from untreated to treated (or vice versa). Other variations on the design may select a treatment group and a similar control group, examining differences over time across the conditions. This version of the difference in difference design was applied by Walcott et al. (2018). However, in this case, the assumption of ignorable treatment assignment is replaced as an assumption of parallel trends. We assume that there is a parallel trend in outcomes in both the treated and untreated groups before the treatment and it would otherwise continue in the absence of treatment.

While fixed effects often observe clusters of units across time, it is not exclusive to longitudinal studies. For example, other researchers have used fixed effects frameworks for twin studies (Oskarsson et al., 2017; Weinschenk et al., 2021), in which the cluster unit is the familial unit. Other studies, leverage clustering of students into different classes such as Figlio et al.'s (2015), use of fixed effects to examine the effects of tenure-track professors on student learning. Such a design could be implemented to examine service-learning outcomes.

Fixed effects regression methods have been used in at least five studies of SLCE, including dissertations. In addition, at least five other studies take a difference-in-differences approach to analysis and use this terminology to describe the analysis. Others apply difference in differences but refer to their work as a pre-post design with nonequivalent control groups, but most do not formally discuss the parallel trends assumption that makes it possible to interpret the results with causal implications. Despite this growing application of the technique, the example from Dahan (2020) is the only publication within a journal focusing on SLCE. This technique can be implemented in a variety of software including SPSS, Stata, R, and SAS.

Propensity Score Analysis

Another technique that can yield plausibly causal effects about SLCE is the use of propensity scores to simulate a counterfactual that enables researchers to compare outcomes among similar treated and untreated units, meeting the ignorable treatment assignment assumption. Conceptually, the propensity to participate in service-learning can be estimated by use of estimation techniques such as logistic regression (Hill et al., 2017). The estimated probability of participation becomes a score that is plausibly independent of the characteristics that otherwise confound the selection of different individuals into the treatment (or not) (Guo & Fraser, 2014). These scores can be used to match individuals with similar propensity scores or may be used as weights in a regression similar to a survey weight that can yield estimates of the difference between the treated and untreated units on an outcome of interest (Hernan & Robins, 2020). Propensity score techniques have software implementations in SPSS, Stata, R, and SAS. Like the difference-in-differences estimator, this technique provides a stronger causal argument for the ATT (vs. ATE), because we must rely on our ability to identify the counterfactual via observed variables.

Recently, the method has been employed to estimate the effects of service-learning participation on underrepresented students' educational success (Song et al., 2017) and sense of belonging (Soria et al., 2019). In addition, the method was used to explore community-engaged student employment as a treatment to examine student retention, graduation, and other indicators of success (Schulzetenberg et al., 2020). The research stemming from the FIPSE grant at University of Minnesota produced several other examples of propensity score methods. This technique has at least a dozen examples including dissertation studies. An expanded discussion of the utility of propensity score analysis for research on community engagement is presented in a special volume of *Journal of Higher Education Outreach and Engagement* (Maruyama et al., 2023).

In these papers, the technique was to match participants and nonparticipants on propensity scores based on set of covariates (often just their demographic information and some precollege variables). By matching on the propensity score, otherwise known as selection on the observables, researchers can offer better estimates of the ATT than simply controlling for these characteristics in traditional regression because the calculated propensity score is theoretically independent of the characteristics and the outcome, blocking the potential for confounding and leading to a defensible ignorable treatment assignment mechanism (Cunningham, 2021).

In particular, the Schulzetenberg et al. (2020) paper has potential implications for high-quality design. In this paper, the authors seek to examine the effect of community-engaged employment opportunities on students from marginalized backgrounds. They condition participation in the activities on a number of characteristics known to contribute to better college outcomes, all of which are observed before students participated (or not) in the employment opportunity during their first year in college. By delimiting the research to only the first-year participants and identifying a sample of students with similar precollege and demographic characteristics with similar propensity scores, the results are particularly compelling. However, the design and conduct of the research were intentional because the researchers partnered with the agency offering the employment opportunity and with the college's institutional research office to access deidentified student records that were on the shelf to conduct this research.

Modeling selection based on the outcome is acceptable to yield a defensible counterfactual, but modeling selection on theoretical treatment selection mechanisms may be better (Guo & Fraser, 2014), but they must be observed before the treatment; otherwise, they produce the same problem that are attempting to solve. The assumption of the propensity score approach is that the selection mechanism is as good as random, conditional on the covariates that model the propensity score. The stronger the relationship between the covariate and the propensity to participate, the better that variable will do to accurately and efficiently provide a propensity score.

Other Techniques for Causal Inference

Two other methods are worth reviewing to encourage their use in research on SLCE: instrumental variables estimation and the synthetic control method. I review each below.

Instrumental Variables Estimation. In the instrumental variables method, a treatment variable features some left-out variable bias in its relationship with an outcome (such as selection bias or confounding). However, this treatment is related to another variable that is independent of that bias; this variable is called an instrument (Angrist & Krueger, 2001). The relationship between that instrument and the desired outcome variable occurs only through the treatment of interest. Conceptually, the instrument causes some of the variation in the treatment of interest and the relationship between the outcome and the part of the variation of the treatment caused by the instrument helps us yield the causal effect of the treatment on the outcome. However, this causal effect is local to the group that is affected by the instrument; therefore, it is called a local average treatment effect (LATE) (Angrist et al., 1996). This technique can be implemented in a wide variety of software including SPSS, Stata, R, and SAS.

In an example of this technique, Mo et al. (2022) employ the selection score for Teach for America (TFA) to estimate how participation in that national service program affects voter turnout among participants that narrowly achieve selection into the program and those that narrowly missed the selection score criteria set by the program. The selection score at the cutoff determines whether a person is invited to participate in TFA. The selection score itself is not independent of the outcome of interest, but close to the cutoff, the individuals are approximately equal in expectation, except that some are invited to participate in a prestigious national service program and others are not. Not everyone, that is invited, chooses to participate; therefore, the jump in participation is not sharp from zero participation to full participation, but the invitation to participate does increase that groups' probability of participation in the program, while those just below the cutoff are much less likely to participate.

This technique is also called a "fuzzy" regression discontinuity. However, fuzzy discontinuities can be valid instrumental variables (Angrist & Pischke, 2008). In these cases, the analysis yields a kind of LATE called the complier's average causal effect (CACE) for the group of individuals induced by the invitation to participate in TFA by narrowly meeting the selection criteria set by the program compared against a group identified as compilers, who did not seek participation when not admitted based on their score.

This design can be employed in a variety of settings in which there is administrative control over programs or activities that may enable valid comparisons across fuzzy discontinuities. For example, universities that offer scholarship programs for intensive community service, such as Bonner Scholar and Leader programs, could rigorously evaluate the programs by comparing the participants to other students who applied but were not admitted to the program if those decisions were made either via a selection score or chosen randomly via a lottery in cases where many students are competing for too few opportunities. The technique could have also been implemented in the analysis of the Moely and Illustre (2011) study of student attitudes to the implementation of the public service requirements at Tulane University, whereby the cohort year of the student would be the instrumental variable.

Synthetic Control Method. Another technique that shows substantial promise for use in SLCE research is the synthetic control group method. This technique is particularly promising because it enables the estimation of plausibly causal effects for single case studies in which the outcome can be observed both before and after the implementation of an intervention. Rather than comparing the outcome in the treated unit against a set of untreated units directly, a set of similar untreated units are identified that can serve as "donor units" to create a synthetic counterfactual of the treated unit (Cunningham, 2021). The technique reweights the observed outcome trend across the donors to mimic the outcome pattern in the treated unit before the intervention as closely as possible, based on a variety of observed covariates (including the pretreatment outcome trend). Using the same donors with the same weights applied, the outcome pattern is plotted in the period of time after the intervention begins in the treated unit. The plots of the outcome pattern in the treated unit and the synthetic counterfactual unit are compared, and we attribute the difference in the trends as being caused by the intervention. This technique has software implementations in both Stata and R.

One of the only applied examples of this technique to SLCE research is Pearl et al. (2013) exploration of the Carnegie Community Engagement Classification on public service spending at land grant institutions, presented at the 2013 Engaged Scholarship Consortium conference. This study used the characteristics such enrollments, expenditures, FTE faculty, and other covariates to form synthetic versions of the five institutions that received the elective classification in 2006 against the set of land grant institutions that had not received the classification by 2010. The results were inconclusive regarding the impact of the classification on public service spending, but the approach was implemented with fidelity.

Because of the hyperlocal nature of engagement activities (Dostilio et al., 2019), identifying potential donor units that do not receive an SLCE intervention is possible by creatively identifying a unit of analysis such as a school district, census tract, or other small area that shares characteristics with other communities but would not be engaged by SLCE activities given their proximity to an engaged university. Fortunately, data for units like school districts and census tracts are publicly available, and there are thousands of potential donors to form the synthetic control. Furthermore, as more institutions collect data using tools like the Collaboratory and GivePulse, it may be possible to identify interventions that may be valid to examine using a data-intensive technique like synthetic counterfactual.

A Path Forward for Causal Research in Service-Learning and Community Engagement

This paper attempts to summarize the variety of approaches to causal inference in research and argues for better design of research on community engagement that yield causal interpretations from research results. The heart of the credibility revolution in econometrics was improvements in research design (Angrist & Pischke, 2010). As demonstrated by the emerging evidence of causal inference strategies applied to SLCE, there are designs that can accommodate the difficulties of managing randomization to treatment and many of these designs would be novel contributions in the SLCE field. Several recommendations stem from this review.

First, researchers must advance stronger arguments for the validity of their findings when applying causal inference strategies. For example, the use of matching and other propensity score techniques offers researchers the opportunity to develop the counterfactual outcome of the treated group. However, even among the small group of exemplars included in this review, the authors were cautious to argue their findings as causal, despite demonstrating strong evidence that their counterfactuals were internally valid comparisons to the treated units. A healthy dose of skepticism is reasonable when making generalization claims about observational data, but when defensible, a causal argument should be advanced. One recommendation for future work applying these methods is to craft stronger arguments around the claims: while the results may not generalize beyond the settings because of the methods, the relative size of the results from the works suggests a plausible causal effect of community engagement.

Second, researchers must consider the mechanisms that produce the causal effects. The core argument of this paper is designed for causal analysis, but articulating the theoretical mechanism that produces the effect is essential for internally valid research. The papers reviewed here provide some guidance for future researchers but more emphasis on the theoretical mechanisms that produce the causal effects is necessary

(Bringle, 2003; Hatcher et al., 2019). As researchers continue to explore the effects of community engagement, articulation of defensible theoretical mechanisms is necessary.

Researchers interested in causal effects must acknowledge that many techniques for causal inference strategies require substantially larger samples than are typically reported in service-learning research. Among the more popularly cited meta-analyses of service-learning research, the average sample size of the included studies was fewer than 140 observations (Burch et al., 2019; Yorio & Ye, 2012) or 200 observations (Celio et al., 2011). The techniques used in the meta-analysis are useful for identifying the overall effect size across many different studies but cannot account for the overestimation of the effect that may occur with uncontrolled selection or other omitted variable bias. However, the idea of collecting many similar measures to identify the effect of an intervention like SLCE may be valuable, especially if a similar approach is implemented and multiple sites collaborate on the research design to yield defensible causal results. Similarly, there remains a need for replication of findings across service-learning research (Adams et al., 2005; Dahan, 2016).

Researchers should consider intentionally collaborating with practitioners to conduct causal studies. SLCE practices are both time and resource intensive, and in hypercompetitive higher education markets, evidence of impact is necessary to justify these investments; as stakeholders in the activities, practitioners need rigorous evidence in support of their work. Practitioners interested in conducting causal research bring many potential advantages for high-quality research design (Clayton et al., 2019). In particular, deep understanding of the mechanisms that may ultimately be useful as instrumental variables or part of the selection on the observables equation to inform how participants in SLCE are sorting into these opportunities. This insider knowledge can be invaluable but may be overlooked without collaboration between practitioners and researchers (Green, 2023). In addition, this insider knowledge lends itself to better measurement of the covariates that may be driving selection, which again improves the implementation of methods like propensity score techniques.

Researchers and practitioners interested in learning more about these methods can access resources related to their implementation via books such as Scott Cunningham's (2021) *Causal Inference: The Mixtape*, and the book's website features software code in Stata and R related to each technique that replicates well-regarded studies in econometrics. In addition, the reference manuals for software like Stata are also useful resources, particularly for the "treatment effects" (propensity score) and "difference-in-differences" implementations (StataCorp, 2021). The software R remains a free, open-source application, with strong documentation and a vast user community. While other software like SPSS and SAS offers users opportunities to implement many designs discussed above, the author is less familiar with these platforms and cannot make good recommendations for their use.

Finally, the SLCE field would benefit from investment in training and development opportunities in causal inference strategies. These investments could take the form of preconference workshops and/or webinars focused on research design and analysis of quantitative data. In addition, as IARSLCE continues to flourish, the formation of special interest groups for research design could emerge as a space for like-minded researchers to collaborate and encourage causal inference techniques to be implemented.

Appendix

Glossary of Terms

Term	Definition
Average Treatment Effect	The treatment effect of interest in true experimental designs, the average treatment effect is the average difference in outcomes for the treated vs the untreated. This result implies high levels of external validity (e.g., generalizability). See also Validity.
Average Treatment Effect on the Treated	The treatment effect of interest in most observational designs, the average treatment effect on the treated is the difference between the treated group and its counterfactual (which are unobserved but estimated by comparing treated outcomes to the outcomes for the untreated group). This treatment effect is justifiable under the assumptions of the potential outcomes model but is limited in its generalizability beyond specific settings. See Counterfactual, Ignorable Treatment Assignment, Exclusion of Potential Outcomes, and Monotonicity.
Complier's Average Causal Effect	Also called a local average treatment effect, the treatment effect of interest in the fuzzy regression discontinuity design and other instrumental variables estimation. The complier subpopulation is defined as the group of individuals who would participate in the treatment when assigned to the treatment and remain untreated when assigned to the control group. The effect is defined as the observed difference in outcomes of the treated group and untreated group divided by the differences in the proportion in the treated group assigned to the treatment and the proportion assigned to the control receiving treatment. For outcome y, treatment w, and assignment z, where $\bar{y}$ is the average outcome and $\bar{w}$ is the proportion receiving treatment: $\frac{(\bar{y}_{w=1,z=1} - \bar{y}_{w=0,z=0})}{(\bar{w}_{z=1} - \bar{w}_{z=0})}$
Counterfactual	A counterfactual can be understood as an alternative state of the factual world. If an action is taken in the factual world, the counterfactual world is one where the action did not take place. Because we cannot observe both the factual world and the counterfactual world, we must approximate the counterfactual world in order to understand the unobserved potential outcome of that world. In this respect, the counterfactual is our attempt to create an internally valid comparison of the outcome that may have happened to compare to the observed outcome of the units that received a treatment. When we can assume that the comparison is based on an ignorable treatment assignment, has potential treated outcomes exclusive to the treated group, and monotonicity in treatment, we have the criteria for to interpret the untreated units as the counterfactual to the treated units, enabling us to estimate the average treatment effect on the treated. See also validity, ignorable treatment assignment, exclusion of potential outcomes, monotonicity, potential outcomes.
Covariate	Any observed or unobserved variable that has covariance with the outcome variable. In selection on the observables, covariates are used to model the selection process into treatments. In the selection on the unobservables, covariates model the selection into unobserved outcomes.
Difference in Differences	Any design that features a comparison of two differences, especially in cases where the first difference is time and the second difference is a change that is exogenous to the outcome (i.e., the treatment). The ignorable treatment assignment assumption takes the form of an assumption of "parallel trends" in the outcomes of treated and untreated units before the treatment is applied. See also parallel trends.
Endogeneity	The condition of dependence between the residual error and the values of the treatment and outcome. Endogeneity can take the form of reverse causality (wherein the dependent variable causing the treatment) or simultaneity (wherein the unobserved variables simultaneously causing the treatment and the outcome). See also Exogeneity, Regression.
Exclusion of Potential Outcomes	The assumption that potential outcomes of any treatment are exclusive to the treated group means that no treated unit influences the potential outcome of the untreated or vice versa. This assumption also means that all treatment effects are represented in the potential outcomes framework and there are no unrepresented versions of the treatment. This assumption is also referred to as the Stable Unit Treatment Value Assumption (SUTVA). This is a fundamental assumption of causal inference. In the case of instrumental variables

	estimation, the instrumental variable cannot otherwise affect the outcome except through the treatment (because it cannot itself be a version of the treatment). See also: Instrumental Variables Estimation, Propensity Score, Monotonicity, Ignorable Treatment Assignment, Potential Outcomes Framework.
Exogeneity	The condition of independence between the treatment and outcome and the residual error. This condition is satisfied in experimental settings by random assignment to conditions. In instances of natural experiments, exogeneity of the treatment assignment must be demonstrated by a discontinuity outside of the control of the participants and the researchers. In other non-experimental settings, the condition of exogeneity may be achieved through selection on the observables, selection on the unobservables, regression discontinuities, or the use of instrumental variables.
Fixed Effects Transformation	The fixed effects transformation is used for linear regression models with clustering of units (such as across time or other clusters like twin pairs) in which the variation within the cluster (e.g. difference in treatment status across time) is retained and all information between clusters is averaged out of the equation because it does not vary over time or otherwise within the cluster. When coupled with a natural experiment, the fixed effect regression removes all omitted variable bias of the observed and unobserved variables across time, leaving only the observable changes in the outcome, the treatment, and time. See also Difference in Differences, Omitted Variable Bias, Exogeneity, Natural Experiment.
Forcing Variable	In regression discontinuity designs, a forcing variable features a discontinuity in the assignment to treatment. For example, a scholarship program uses a cutoff selection score in making offers. If the selection score completely determines who receives and who does not receive the scholarship, the regression discontinuity is said to be sharp. If the selection score is substantially informative about who receives and does not receive scholarships (but not perfect), the regression discontinuity is said to be fuzzy. However, based on the value of the forcing variable, individuals on both sides of the discontinuity are approximately equal in expectation under the ignorable treatment assignment assumption. See also Ignorable Treatment Assignment, complier's average causal effect.
Fuzzy Regression Discontinuity Design	This design features many of the same design features as the sharp regression discontinuity design but treats the cutoff value in the forcing variable as an instrumental variable to estimate the complier's average causal effect. See also Sharp Regression Discontinuity Design, complier's average causal effect
Ignorable Treatment Assignment	Treatment assignment is ignorable when, conditional on covariates, potential outcomes for treatment ($W = 1$) are independent of potential outcomes for untreated ($W = 0$). The assumption is credible when the observable characteristics of the treated and untreated samples are balanced and equal, suggesting the potential outcomes are balanced and equal in expectation. Through propensity score methods, this is achieved when the propensity score is used to balance the samples. This is a fundamental assumption of the potential outcomes framework. See also Propensity Score, Selection on the Observables, Covariate, Forcing Variable, Regression Discontinuity Design.
Instrumental Variables Estimation	An instrument is a variable that (1) is related to a treatment that is endogenous to an outcome, (2) is itself exogenous to that outcome (i.e. ignorable treatment assignment), and (3) only causes variation between the outcome and itself through the endogenous treatment (exclusive potential treated outcomes and monotonicity). For example, the distance a person lives from the nearest college has been used multiple times as an instrumental variable to demonstrate the effect of additional education on a variety of outcomes because the distance to the nearest college is "as good as random" relative to people's place of residence, but this distance may influence the decision to pursue a college education because the cost to attend is lower for a person that lives nearby a college. The only way that distance would affect outcomes that are thought to be influenced by educational attainment is through the increases in attainment attributable to the proximity of the college, which satisfies the exclusive treatment assumption (n.b. for a full example of distance to college as an instrumental variable, see Dee, *Are there civic returns to education?* (2004)). In cases where there are fuzzy discontinuities in the treatment status by individuals based on the value of some forcing variable, the location of the discontinuity is a valid instrumental variable to

	identify the local average treatment effect. See complier's average causal effect, Fuzzy Regression Discontinuity Design, Monotonicity, Exclusion of Potential Outcomes.
Intent to Treat Effect	A conservative estimate of the treatment effect in settings where there is imperfect compliance, the intent to treat ignores the actual treatment participation and compares the intended treatment assignment. This effect is valid under random assignment because all participants have equal expectation regardless of their actual participation status, but in non-experimental settings (like fuzzy regression discontinuities) this effect is confounded by the uncontrolled selection, but the complier's average causal effect recovers the treatment effect among the subpopulation of compilers. See also Random Assignment, Fuzzy Regression Discontinuity, complier's average causal effect.
Monotonicity	Assignment to treatment W is conditional on the value of an assignment variable Z such that $W(Z) = W(1) \geq W(0)$. In a compliance framework, this assumption means there are no "defiers" who would refuse treatment when assigned to the treatment group but receive treatment when assigned to the control group. The assumption that this group does not exist is a fundamental assumption for causal inference. See also: instrumental variables estimation, complier's average causal effect.
Natural Experiment	Any design that leverages natural discontinuities in treatment status that are outside of the control of the units affected by the treatment. Commonly, policy changes due to federalism offer plausible mechanisms to compare the treatment effects of policy implementations, but natural disasters can also serve as valid discontinuities that can be used to estimate treatment effects. See also difference in differences, exogeneity.
Omitted Variable Bias	The omitted variable bias can be understood to be the result of a missing value problem. In any study, the observed potential outcome for untreated units is only the untreated potential outcome, while the treated potential outcome is observed for the treated units. However, when the units' selection into treatments is uncontrolled, we cannot understand the potential treated outcome among the untreated units because their outcome is confounded by their selection into the treatment. In the absence of some characteristic that informs the selection into treatments, the observational difference for each unit contains both the observed and *corresponding unobserved* value for a treated and untreated potential outcome. The direction of the omitted variable's bias (in the form of the extra potential outcomes) is unknown. Formally, omitted variable bias can be seen in the following equation for uncontrolled selection into treatment W for units i, and outcome Y: $$Y_i = \begin{cases} Y_{1i} if\ W_i = 1 \\ Y_{0i}\ if\ W_i = 0 \end{cases}$$ $$Y_{0i} + (Y_{1i} - Y_{0i})W_{\text{uncontrolled}}$$ $$= E[Y_i \mid W_{\text{uncontrolled}} = 1] - E[Y_i \mid W_{\text{uncontrolled}} = 0]$$ $$= \text{observed}(E[Y_{1i} \mid W_i = 1] - E[Y_{0i} \mid W_i = 0]) + \text{unobserved}(E[Y_{0i} \mid W_i = 1] - E[Y_{1i} \mid W_i = 0])$$
Parallel Trends	An alternative form of the ignorable treatment assignment assumption applied to difference-in-differences analysis in which the values of outcome of the treated and untreated groups may be unequal, but their trend over time is roughly equal. The assumption of parallel trends enables the change in treatment status from untreated to treated causes the change in the outcome trend for the treated group, while the outcome trend in the untreated group is interpreted as the likely counterfactual trend for the treated units in the absence of the treatment. See also Counterfactual, Difference in Differences, Ignorable Treatment Assignment.
Potential outcomes	The potential outcomes framework uses a counterfactual logic to understand the treatment effect of an intervention W on an outcome Y. At its core, the logic relies on three assumptions about the counterfactual that enable the comparison of any treated and untreated units for causal inference: ignorable treatment assignment, exclusion of potential outcomes, and monotonicity. The potential outcomes framework uses the following equation to estimate the outcome and this is further decomposed into an average treatment effect:

	$$Y_i = \begin{cases} Y_{1i} if\ W_i = 1 \\ Y_{0i}\ if\ W_i = 0 \end{cases}$$ $$= Y_{0i} + (Y_{1i} - Y_{0i})W_i$$ $$= E[Y_i \mid W_i = 1] - E[Y_i \mid W_i = 0]$$ See also: Ignorable Treatment Assignment, Exclusion of Potential Outcomes, Monotonicity.
Propensity Score	The calculated probability of treatment assignment generated from observed covariates measured before a treatment is applied or otherwise exogenous covariates. Under the ignorable treatment assignment assumption, the propensity score is conditionally independent of the covariates that are used to calculate it and is conditionally independent of the observed treatment status of any unit. With this score, various propensity score methods can be implemented such as optimal matching techniques or inverse probability of treatment weighting techniques. See also ignorable treatment assignment, covariate, selection on the observables.
Quasi-Experiment	Any non-experimental design that non-randomly assigns units to treatment and control. See also difference in differences, regression discontinuity design, natural experiment, and propensity score.
Random Assignment	Assignment to treatment W is randomly determined, often with equal probability of treatment and control. When randomly assigned to treatment and control, all units have equal expectation in their outcome means, so any differences in the means is attributable to the treatment effect. See Average Treatment Effect
Regression	A set of statistical techniques that model the covariances between multiple predictor variables (called covariates) and one or more outcomes of interest. In linear regression, a continuous outcome is predicted with a line based on the observed values of the covariates, with differences from the predicted line being called residual error. The technique called ordinary least squares most efficiently minimizes the residual error for linear regression, but may extrapolate the outcome beyond the scope of the data and also assumes that the observed covariates are independent of the residual error. Any unobserved variable is captured in the residual error and their relationship with the outcome or the covariates may influence the coefficients for the covariates, leading to omitted variable bias. Other regressions include logistic regression for binary and “multinomial” outcomes, poisson regression for count outcomes, and Cox regressions for time to event outcomes. See also Endogeneity, Fixed Effects Transformation, Omitted Variable Bias.
Regression Discontinuity Design	This design features a discontinuity in treatment assignment related to a forcing variable. The discontinuity either completely or partially determines treatment status but is otherwise exogenous to the outcome of interest and other covariates. The difference in the outcome by treatment status is valid near to the discontinuity, yielding a local average treatment effect.
Selection on the Observables	In response to the omitted variable bias problem, observable characteristics inform the selection into treated and untreated conditions. Using the propensity to select into treatment by conditioning on the observable characteristics, we block the additional confounding of the unobserved potential outcome to the equation and achieve a defensible assumption of ignorable treatment assignment. See also Omitted Variable Bias, Selection on the Unobservables, Potential Outcomes, Ignorable Treatment Assignment.
Sharp Regression Discontinuity Design	This design features a sharp discontinuity in treatment assignment at a cutoff value of a forcing variable, such that the treatment status jumps from 0 to 1 across the cutoff for all units. While the effect is a complier’s average causal effect, there are no units that did not comply with their treatment assignment, so no adjustments are made and the applicability of a finding is not limited to a subset of the observed population, but to all observed units.
Synthetic control	Method for estimating the untreated potential outcome of a single aggregate unit by creating a “synthetic” control unit that mimics the observable characteristics of the treated unit during the pre-treatment period. The synthetic unit is a composite of weighted values of “donor units”, similar to those calculated and applied in a propensity score weighting method. When compared in the post-treatment period, the average difference in the outcome trends for the treated unit and its synthetic control unit is understood to be the causal effect

	of the treatment in the treated unit. See also Counterfactual, Difference in Differences, Propensity Score.
Validity	Donald Campbell classified four types of validity: statistical conclusion validity as inferences about the covariation between a treatment and outcome; construct validity as the operationalization of constructs intended to represent theoretical relationships; internal validity as the ability to attribute the relationship between treatment and outcome as causal; and external validity as the ability to hold the conclusion of the cause-effect relationship as generalizable beyond the units, settings, treatments, and outcomes observed (Shadish, 2010). While each kind of validity is valuable to research, much of the discussion contained within this article regarding validity concerns internal validity.

Acknowledgement

I would like to acknowledge the feedback and encouragement of Andrew Pearl, Robert Bringle, Richard Blissett, and others who discussed this manuscript with me at the IARSLCE 2023 and AERA 2024 conferences.

References

Adams, A. T., Ajrouch, K. J., Henderson, H., & Heard, I. (2005). Service-learning outcomes research: The role and scarcity of replication studies. *Journal of Applied Sociology, 22*(2), 55–74. https://doi.org/10.1177/19367244052200204

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Angrist, J. D., & Krueger, A. B. (2001). Instrumental variables and the search for identification: From supply and demand to natural experiments. *The Journal of Economic Perspectives, 15*(4), 69–85. https://doi.org/10.1257/jep.15.4.69

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Angrist, J. D., & Pischke, J.-S. (2010). The credibility revolution in empirical economics: How better research design is taking the con out of econometrics. *Journal of Economic Perspectives, 24*(2), 3–30. https://doi.org/10.1257/jep.24.2.3

Blalock, H. M. (2018). *Causal inferences in nonexperimental research*. UNC Press Books.

Bowman, N. A., Park, J. J., & Denson, N. (2015). Student involvement in ethnic student organizations: Examining civic outcomes 6 years after graduation. *Research in Higher Education, 56*(2), 127–145. https://doi.org/10.1007/s11162-014-9353-8

Bringle, R. G. (2003). Enhancing theory-based research on service-learning. *Deconstructing Service-Learning: Research Exploring Context, Participation, and Impacts, 3*(2), 3–21.

Brown, M. A. (2011). Learning from service: The effect of helping on helpers' social dominance orientation. *Journal of Applied Social Psychology, 41*(4), 850–871. https://doi.org/10.1111/j.1559-1816.2011.00738.x

Burch, G. F., Giambatista, R., Batchelor, J. H., Burch, J. J., Hoover, J. D., & Heller, N. A. (2019). A meta-analysis of the relationship between experiential learning and learning outcomes. *Decision Sciences Journal of Innovative Education, 17*(3), 239–273. https://doi.org/10.1111/dsji.12188

Campbell, D. T., & Cook, T. D. (1979). *Quasi-experimentation*. Rand Mc-Nally.

Celio, C. I., Durlak, J., & Dymnicki, A. (2011). A meta-analysis of the impact of service-learning on students. *Journal of Experiential Education, 34*(2), 164–181. https://doi.org/10.1177/105382591103400205

Clayton, P. H., Stokamer, S., Garvin, L., Shoemaker, D., Muse, S., & Norvell, K. (2019). Practical wisdom on co-inquiry in research on service learning. In J. A. Hatcher, R. G. Bringle, & T. W. Hahn (Eds.), *Practical wisdom for conducting research on service learning: Pursuing quality and purpose*. Stylus Publishing.

Cunningham, S. (2021). *Causal inference: The mixtape*. Yale University Press.

Dahan, T. A. (2016). Revisiting pedagogical variations in service-learning and student outcomes. *International Journal of Research on Service-Learning and Community Engagement, 4*(1), 3–15. https://doi.org/10.37333/001c.29586

Dahan, T. A. (2020). Did federal policy on postsecondary service-learning support community social capital? *Journal of Higher Education Outreach and Engagement, 24*(3), 3–20.

Danley, S., Perry, A., McNeese, J., & Razi, L. (2022). Ethnography without experimentation: Ethnographic methods in post-disaster contexts. In J. D. Rivera (Ed.), *Disaster and emergency management methods: Social science approaches in application* (1st ed.). Routledge. https://doi.org/10.4324/9780367823948-13

Dee, T. S. (2004). Are there civic returns to education? *Journal of Public Economics, 88*(9–10), 1697–1720. https://doi.org/10.1016/j.jpubeco.2003.11.002

Denzin, N. K. (2008). The new paradigm dialogs and qualitative inquiry. *International Journal of Qualitative Studies in Education, 21*(4), 315–325. https://doi.org/10.1080/09518390802136995

Dostilio, L. D., Ohmer, M. L., McFadden, K., Mathew, S., & Finkelstein, C. (2019). *Building community capacity via place-based hyperlocal engagement*. Coalition for Urban and Metropolitan Universities. https://www.cumuonline.org/research/hyperlocal-research/

Figlio, D. N., Schapiro, M. O., & Soter, K. B. (2015). Are tenure track professors better teachers? *Review of Economics and Statistics, 97*(4), 715–724. https://doi.org/10.1162/REST_a_00529

Goodkind, J. R., Amer, S., Christian, C., Hess, J. M., Bybee, D., Isakson, B. L., Baca, B., Ndayisenga, M., Greene, R. N., & Shantzek, C. (2017). Challenges and innovations in a community-based participatory randomized controlled trial. *Health Education & Behavior, 44*(1), 123–130. https://doi.org/10.1177/1090198116639243

Green, P. M. (2023). The Scholar-Administrator Imperative: Developing scholarship and research through practice to build the community engagement field. *Metropolitan Universities, 34*(3), Article 3. https://doi.org/10.18060/26863

Guo, S., & Fraser, M. W. (2014). *Propensity score analysis: Statistical methods and applications*. SAGE.

Hatcher, J. A., Bringle, R. G., & Hahn, T. W. (2019). *Practical wisdom for conducting research on service learning: Pursuing quality and purpose*. Stylus Publishing.

Hempel, C. G. (1950). Problems and changes in the empiricist criterion of meaning. *Revue Internationale de Philosophie, 4*(11), 41–63.

Hernan, M. A., & Robins, J. M. (2020). *Causal inference: What if.* Chapman & Hall. https://www.hsph.harvard.edu/miguel-hernan/wp-content/uploads/sites/1268/2024/01/hernanrobins_WhatIf_2jan24.pdf

Hill, P. L., Pasquesi, K., Bowman, N. A., & Brandenberger, J. W. (2017). Longitudinal research and student civic outcomes. In J. A. Hatcher, R. G. Bringle, & T. W. Hahn (Eds.), *Research on student civic outcomes in service learning* (pp. 393–424). Routledge.

Hume, D. (2007). *A treatise of human nature: Texts* (Vol. 1). Clarendon Press.

Imbens, G. W., & Rubin, D. B. (2010). Rubin causal model. In S. N. Durlauf & L. E. Blume (Eds.), *Microeconometrics* (pp. 229–241). Palgrave Macmillan UK. https://doi.org/10.1057/9780230280816_28

Jamieson, M. K., Govaart, G. H., & Pownall, M. (2023). Reflexivity in quantitative research: A rationale and beginner's guide. *Social and Personality Psychology Compass, 17*(4), e12735. https://doi.org/10.1111/spc3.12735

Jones, S. R., & Foste, Z. (2016). Qualitative research on service learning and student civic outcomes. In J. A. Hatcher, R. G. Bringle, & T. W. Hahn (Eds.), *Research on student civic outcomes in service learning: Conceptual frameworks and methods* (Vol. 3, pp. 238–251). Stylus Publishing.

Leamer, E. E. (1983). Let's take the con out of econometrics. *The American Economic Review, 73*(1), 31–43.

Light, R. J., Singer, J. D., & Willett, J. B. (2009). *By design: Planning research on higher education*. Harvard University Press.

Markus, G. B., Howard, J. P. F., & King, D. C. (1993). Notes: Integrating community service and classroom instruction enhances learning: Results from an experiment. *Educational Evaluation and Policy Analysis, 15*(4), 410–419. https://doi.org/10.3102/01623737015004410

Maruyama, G., Furco, A., & Song, W. (2018). Enhancing underrepresented students' success through participation in community engagement. In T. D. Mitchell & K. M. Soria (Eds.), *Educating for citizenship and social justice: Practices for community engagement at research universities* (pp. 221–235). Palgrave Macmillan.

Maruyama, G., Lopez, I., Schulzetenberg, A., & Song, W. (2023). Why researchers should consider using propensity score matching methods to examine effectiveness of community engagement programming. *Journal of Higher Education Outreach and Engagement, 27*(2), Article 2. https://openjournals.libs.uga.edu/jheoe/article/view/3040

Maxwell, J. A. (2012). The importance of qualitative research for causal explanation in education. *Qualitative Inquiry, 18*(8), 655–661. https://doi.org/10.1177/1077800412452856

Mitchell, T. D., & Rost-Banik, C. (2017). Critical theories and student civic outcomes. In J. A. Hatcher, R. G. Bringle, & T. W. Hahn (Eds.), *Research on student civic outcomes in service learning: Conceptual frameworks and methods* (pp. 169–189). Stylus Publishing.

Mo, C. H., Holbein, J. B., & Elder, E. M. (2022). Civilian national service programs can powerfully increase youth voter turnout. *Proceedings of the National Academy of Sciences, 119*(29), e2122996119. https://doi.org/10.1073/pnas.2122996119

Moely, B. E., & Ilustre, V. (2011). University students' views of a public service graduation requirement. *Michigan Journal of Community Service Learning, 17*(2), 43–58.

Morgan, S. L., & Winship, C. (2015). *Counterfactuals and causal inference*. Cambridge University Press.

Murnane, R. J., & Willett, J. B. (2010). *Methods matter: Improving causal inference in educational and social science research*. Oxford University Press.

Osborne, R. E., Hammerich, S., & Hensley, C. (1998). Student effects of service-learning: Tracking change across a semester. *Michigan Journal of Community Service Learning, 5*, 5–13.

Oskarsson, S., Thisted Dinesen, P., Dawes, C. T., Johannesson, M., & K. E. Magnusson, P. (2017). Education and social trust: Testing a causal hypothesis using the discordant twin design. *Political Psychology, 38*(3), 515–531. https://doi.org/10.1111/pops.12343

Pearl, A., Byars, J., & Turk, J. (2013). *Impact of Carnegie's community engagement classification: A synthetic control approach*. Engagement Scholarship Consortium.

Pong, H.-K., & Lam, P. (2023). The effect of service learning on the development of trait emotional intelligence and adversity quotient in youths: An experimental study. *International Journal of Environmental Research and Public Health, 20*(6), Article 6. https://doi.org/10.3390/ijerph20064677

Popper, K. R. (1979). *Objective knowledge: An evolutionary approach* (Vol. 49). Clarendon Press.

Schulzetenberg, A. J., Wang, Y.-C., Hufnagle, A. S., Soria, K. M., Maruyama, G., & Johnson, J. (2020). Improving outcomes of underrepresented college students through community-engaged employment. *International Journal of Research on Service-Learning and Community Engagement, 8*(1). https://doi.org/10.37333/001c.18719

Scriven, M. (2008). A summative evaluation of RCT methodology & an alternative approach to causal research. *Journal of MultiDisciplinary Evaluation, 5*(9), Article 9. https://doi.org/10.56645/jmde.v5i9.160

Shadish, W. R. (2010). Campbell and Rubin: A primer and comparison of their approaches to causal inference in field settings. *Psychological Methods, 15*(1), 3–17. https://doi.org/10.1037/a0015916

Shumer, R. (2000). Science or storytelling: How should we conduct and report service-learning research? *Michigan Journal of Community Service Learning, S-1*, 76–83.

Song, W., Furco, A., Lopez, I., & Maruyama, G. (2017). Examining the relationship between service-learning participation and the educational success of underrepresented students. *Michigan Journal of Community Service Learning, 24*(1), 23–37. http://doi.org/10.3998/mjcsloa.3239521.0024.103

Soria, K. M., Hufnagle, A. S., & Do, T. (2019). Exploring the differential effects of service-learning on students' sense of belonging: Does social class matter? *International Journal of Research on Service-Learning and Community Engagement, 7*(1). https://doi.org/10.37333/001c.11486

StataCorp. (2021). *Stata MP* (Version 17.0) [Computer software].

Steinberg, K. S., Bringle, R. G., & McGuire, L. E. (2013). Attributes of high-quality research on service learning. *Research on Service Learning: Conceptual Frameworks and Assessment, 2*, 27–53.

Walcott, R. L., Murcia, A. M., Berry, G. M., Juna, C. F., Roldós, M. I., & Corso, P. S. (2018). The impact of nursing students on the health-related quality of life and perceived social support of a rural population in Ecuador: Effects of a service-based learning course. *International Journal for Equity in Health, 17*(1), 16. https://doi.org/10.1186/s12939-018-0734-z

Weinschenk, A. C., Dawes, C. T., & Oskarsson, S. (2021). Does education instill civic duty? Evidence from Monozygotic Twins in the United States and Sweden. *International Journal of Public Opinion Research, 33*(1), 183–195. https://doi.org/10.1093/ijpor/edaa006

Yorio, P. L., & Ye, F. (2012). A meta-analysis on the effects of service-learning on the social, personal, and cognitive outcomes of learning. *Academy of Management Learning & Education, 11*(1), 9–27. https://doi.org/10.5465/amle.2010.0072

About the Author

Thomas A. Dahan, Ph.D. https://orcid.org/0000-0001-7292-9845, is Assistant Teaching Professor in the Rutgers School of Nursing–Camden and Director of the Rutgers Camden Public Health AmeriCorps. Correspondence concerning this article should be addressed to Thomas A. Dahan through email at tom.dahan@camden.rutgers.edu.