

A Meta-analysis of the Impact of Service-Learning on Students

Christine I. Celio, Joseph Durlak, and Allison Dymnicki

Service-learning (SL) has become a popular teaching method everywhere from elementary schools to colleges. Despite the increased presence of SL in the education world, it is still unclear what student outcomes are associated with SL programs and what factors are related to more effective programs. A meta-analysis of 62 studies involving 11,837 students indicated that, compared to controls, students participating in SL programs demonstrated significant gains in five outcome areas: attitudes toward self, attitudes toward school and learning, civic engagement, social skills, and academic performance. Mean effects ranged from 0.27 to 0.43. Furthermore, as predicted, there was empirical support for the position that following certain recommended practices—such as linking to curriculum, voice, community involvement, and reflection—was associated with better outcomes. Current data should be gratifying for educators who incorporate SL into their courses, and should encourage more SL research to understand how students benefit and what conditions foster their growth and development.

Keywords: Service-Learning, Civic Engagement, Meta-analysis

Christine I. Celio, Ph.D., is a clinical psychology fellow at the San Francisco VA Medical Center, California, USA. E-mail: ccelio@luc.edu

Joseph Durlak, Ph.D., is an Emeritus Professor at Loyola University Chicago, Illinois, USA. E-mail: jdurlak@luc.edu

Allison Dymnicki, Ph.D., is a researcher at the American Institutes of Research in Washington, D.C., USA. E-mail: adymni2@uic.edu

Service-learning (SL), defined as a teaching and learning strategy that attempts to integrate community service with an academic curriculum, has become widespread in the United States (Learn and Serve America, 2010). In 1999, 32% of all public schools created service opportunities as part of their curriculum, including nearly half of all high schools (Skinner & Chapman, 1999). Many community colleges and four-year universities also offer service programs. Further exemplifying the popularity of service-learning, Campus Compact, created by the presidents of Brown, Georgetown, and Stanford universities, was designed to infuse service and civic engagement into college academic programs. It started with just a handful of schools interested in service in 1987 and now boasts more than 1,100 schools (Campus Compact, 2009). SL programs usually have a positive influence on the community receiving services, on the educational institution hosting the program (through enhanced and more engaging curriculum offerings), and, finally, on the student participants who may benefit personally, socially, or academically (e.g., Billig, 2009; Conway, Amel, & Gerwien, 2009; White, 2001). The focus of this paper is on the latter area, which has received the most attention from researchers. Despite the growing popularity of service-learning, it is still unclear what student outcomes are associated with SL programs and what factors are related to more effective programs. This meta-analysis focuses on these two research issues.

Student Outcomes

Several research studies suggest that student participation in SL is associated with positive outcomes in five areas: attitudes toward self, attitudes toward school and learning, civic engagement, social skills, and academic achievement (e.g., Billig, 2009; Conway et al., 2009; White, 2001). For example, SL students have demonstrated increases in self-esteem and self-concept, more highly internalized moral standards, more positive attitudes toward school and education, greater interest in, commitment to, and sensitivity toward their communities and their needs, and stronger beliefs that one can make a difference in the world (Billig, Root, & Jesse, 2005). SL students have also grown in various social skills related to communication, leadership, and problem solving. Finally, SL can also lead to improved academic achievement (e.g., Billig, 2009; Giles & Eyler, 1994; Harwood & Radoff, 2009; Markus, Howard, & King, 1993). At the same time, findings in each of the aforementioned areas have not been consistent, as some research has failed to obtain significant effects in these areas

(e.g., Astin, Vogelgesang, Misa, Anderson, Denson, Jayakumar, Saenz, & Yamamura, 2006; Billig et al., 2005; Blyth, Saito, & Berkas, 1997; Eyler, Giles, Stenson, & Gray, 2001; Larkin & Mahoney, 2006; Parker-Gwin & Mabry, 1998).

Two meta-analyses have concluded that SL programs do lead to positive gains for students in multiple outcome areas (Conway et al., 2009; White, 2001). However, White's review only included 12 quantitative studies, most of which did not have a control group. Conway et al. examined a much larger sample of 103 interventions, but they included studies of community service or volunteerism as well as SL projects, and many of these studies did not have control groups. Because of the many threats to validity contained in one-group studies, there is a need for a meta-analysis of controlled outcome studies. Outcomes can be affected by factors such as variations in the methodology and the participants' educational level (elementary, high school, or college), so we also sought to examine how these features might moderate outcomes.

Does Including Recommended Practices Strengthen Program Outcomes?

In understanding the benefits of SL programs, it is also important to determine if the inclusion of recommended practices leads to stronger effects. In 1998, the National Service-Learning Cooperative created an initial list of 11 essential elements of service-learning (National Service-Learning Cooperative, 1998). In 2008, this list was revised, updated, divided into eight sections, and called the *K-12 Service-Learning Standards for Quality Practice* (National Youth Leadership Council, 2011). Limited information in many reviewed studies prevented us from analyzing the influence of all eight standards, but we were able to focus on four of them. These four relate to (a) linking programs to academic and program curriculum or objectives, (b) incorporating youth voice, (c) involving community partners, and (d) providing opportunities for reflection. These standards are consistent with the views of many leading proponents of SL regarding key components of effective SL programs (Billig, 2009). A brief discussion of the rationale for these standards, which we call *recommended practices*, follows.

First, SL programs should be aligned with academic and programmatic curricula, outline clear program goals and objectives for the students, and link the program activities to these goals. Establishing clear goals for students and making explicit connections between service and learning has been linked in some studies to stronger student academic engagement and performance outcomes (Billig et al., 2005), larger increases in problem-solving skills (Conrad & Hedin, 1982), and improved learning and satisfaction with the program (Hamilton & Zeldin, 1987). Providing opportuni-

ties for students to transfer or apply what they have learned to multiple contexts has also been associated with improved learning outcomes (Boss, 1994; National Research Council, 1999).

Second, educators should provide youth with a strong voice in planning, implementing, and evaluating SL experiences with guidance from adults. Some data suggest that students who choose what issues to address in their SL projects make greater gains in civic knowledge (Billig et al., 2005). Furthermore, engagement in service-learning has been a strong predictor of other positive outcomes, such as improving self-efficacy, becoming attached to school and community, valuing academics, and becoming more civically engaged in general (Melchior & Bailis, 2002; Meyer, 2006; Perry & Katula, 2001). Youth voice has also been shown to predict the impact of SL in a range of important domains, such as attitudes toward out-groups (Morgan & Streb, 2001).

Third, because one SL goal is to improve or help the community, educators should develop community partnerships and solicit and accept community input on the desired elements and goals of service projects. It is assumed that meaningful SL involves service that strengthens community ties and forms positive relationships, meets some of the community's needs, and, ideally, benefits both the community partners and the students (Gray et al., 1998; Harwood & Radoff, 2009). Researchers have found that strong community partnerships are associated with long-term program sustainability (Ammon, Furco, Chi, & Middaugh, 2002; Billig, 2002; Kramer, 2000).

Another standard emphasizes that opportunities for reflection are critical in order to provide the transformative link between the action of "serving" and the ideas of "learning" (Billig, 2009; Eyler, Giles, & Schmiede, 1996). Some studies have found that reflection is associated with students' experiencing increased self-confidence and engagement in school, greater civic knowledge and social responsibility, and more caring relationships with others (e.g., Anderson, 1998; Billig et al., 2005; Blyth et al., 1997). Although the National Youth Leadership Council's (2011) promulgation of SL standards assumes that including these best practices leads to more effective SL programs, there is a need to examine the empirical support for this assumption.

In sum, this review sought to examine the benefits to participants of SL programs and to assess variables that might moderate student outcomes. First, we hypothesized that SL programs would yield significant positive effects in five student domains: attitudes toward self, attitudes toward school and learning, civic engagement, social skills, and academic performance. Second, we hypothesized that programs that included recommended practices would achieve significantly higher mean effects than those not following recommended practices.

Methodology

Literature Search

Four methods were used to locate relevant studies. First, using the search words “service-learning,” “community service,” “experiential learning,” “public service,” “civic engagement,” and “civic involvement,” a literature search was performed in PsycINFO and ERIC computer databases. Second, to find relevant studies between the years of January 1, 1970, through April 1, 2008, a manual search was conducted in the following journals: *American Journal of Community Psychology*, *College Student Journal*, *Journal of Experiential Education*, *Journal of Adolescence*, *Journal of Early Adolescence*, *Journal of Prevention and Intervention in the Community*, *Michigan Journal of Community Service Learning*, and *Teaching of Psychology*. Third, reference lists from all included studies and from SL books were inspected, as was the annotated bibliography on the impacts of service-learning by Eyler et al. (2001). Fourth, some leading experts of the SL and civic engagement community and many senior scholars attending the 2007 Service-Learning Emerging Scholars meeting were contacted and asked for recommendations of studies to examine.¹

Inclusion Criteria

To be included in the review, the studies had to meet six criteria: (a) appear in English before April 1, 2008; (b) evaluate a SL program that fits the definition of service-learning as an intervention that attempts to integrate service with an academic curriculum; (c) involve students at the elementary, secondary, or postsecondary level; (d) use a control group; (e) contain sufficient information to calculate effect sizes; and (f) evaluate the SL course as the sole primary program component (i.e., studies in which SL was part of a larger multicomponent intervention were not included). All types of published and unpublished reports were eligible for inclusion.

Sixty-three separate programs described in 61 reports were originally considered for this meta-analysis. One paper was removed from analyses because all effects were extreme outliers; therefore, the final count of studies was 62.²

¹ The authors wish to thank the following people for their input and suggestions: Janet Eyler (Vanderbilt University), Chris Chapman (National Center for Educational Studies), Constance Flanagan (Penn State University), Deborah Hecht (New York University), Mark Hugo Lopez (CIRCLE/University of Maryland), Julie C. Rodriguez (César E. Chávez Foundation), Rob Shumer (University of Minnesota), Judith Torney-Purta (University of Maryland), Wendy Wheeler (Innovation Center for Youth and Community), Larry Bailis (Brandeis University), Andy Furco (University of Minnesota), Peter Levine (CIRCLE), and Alan Melchior (Brandeis University).

² The study removed from the final analysis produced effects that were more than four standard deviations beyond the original mean of 0.49 (Leming, 2001). That is, this study produced six effects ranging from -4.13 to 14.6 and was considered to be an aberrant finding.

Coding Procedure

A coding manual was created to identify publication, participant, program, methodological, and outcome features of each study.

Publication features. These included author, title, year of report, and source of report (i.e., whether it was published or unpublished).

Participant features. The grade level of participants was coded into one of five categories: elementary (K–Grade 5), middle (Grades 6–8), high school (Grades 9–12), college undergraduate, and professional school (e.g., students in postbaccalaureate nursing, medical, or social work schools). The predominant ethnicity of the participants (i.e., > 50%) was coded as Caucasian, African American, Latino, Asian-American, and Native American. The predominant socio-economic status (SES) of participants (> 50%) was coded as lower/working class, middle/upper class, or mixed (i.e., at least 20% of participants were drawn from two of the listed groups).

Program features. The presence or absence of each of four recommended SL practices was coded dichotomously as “yes” or “no”: (a) linking to curriculum, (b) youth voice, (c) community involvement, and (d) reflection. To be counted as *linking to curriculum*, the study must have reported, at a minimum, having clear goals for the program that align with the curriculum, and containing corresponding activities to match those goals. *Youth voice* was coded when students were involved in the planning, implementation, or evaluation process of the program. *Community involvement* was coded if the study mentioned that the community had some part in the program besides providing a place for students to serve (e.g., the community was surveyed about their needs during the planning process or about the program’s impact during the evaluation process). To be coded positively for *reflection*, the studies needed to mention some type of reflection activity (e.g., using journals, having discussions in class or in small groups, writing essays about the service experience, presenting to the class what was learned, or reflecting individually with the teacher or site supervisor).

Methodological features. Five methodological features were also examined, including if the study (a) employed a randomized experimental design, (b) conducted assessments at both pre and post, (c) reported using a reliable outcome measure, (d) reported using a valid measure, and (e) identified the source of outcome data (e.g., student, observer, or school record data). The first four of these variables were coded dichotomously.³

An outcome measure’s reliability was recorded if the authors reported that the measure was “reliable,” if the measure was of known

³ “No” was recorded if the author did not provide information on the methodological feature in question and this information could not be tracked down in another related report.

reliability (e.g., the Rosenberg Self-Esteem Scale), or if the authors reported numerically the reliability of the measure (Durlak, Weissberg, Dymnicki, Taylor, & Schellinger, 2011). A measure was considered valid if it was a measure of known validity or if the authors mentioned the measure's construct, concurrent, or predictive validity. The source of outcome data was coded as student (self-report), observer (independent rating), or school record data (data taken from school records).

Outcome features. Study outcomes were organized into five main categories: (a) attitudes toward self, (b) attitudes toward school and learning, (c) civic engagement, (d) social skills, and (e) academic achievement. *Attitudes toward self* included measures related to self-esteem, self-efficacy, personal abilities, and feelings of control. *Attitudes toward school and learning* were defined as students' feelings about school or class (e.g., academic engagement, or enjoyment of the course). *Civic engagement* included any outcome measure oriented toward, or directly affecting, the community (i.e., altruism, civic responsibility, and current and future voting behaviors). *Social skills* included skills generally directed toward other people, such as leadership skills, cultural competence, and social problem solving. *Academic achievement* included measures of students' grades or test performance.

Index of Effect Size and Statistical Procedures

Effect size (ES) assesses the magnitude or strength of the findings that occur in research studies. Effect sizes were calculated as a standardized mean difference in which the post mean of the control group was subtracted from the post mean of the service group and divided by the pooled standard deviation (*SD*) of the two groups. If the group had a pre-ES, the pre-ES was calculated in a similar fashion and then subtracted from the post-ES to determine the overall ES (Hedges & Olkin, 1985). In all cases, positive ES values indicate the service group was superior to the controls at post. When means and standard deviations were not available, methods described by Lipsey and Wilson (2001) were used. When calculating effect sizes for outcomes in which "no effect" or "no significant effect" was reported, we followed the common practice of assigning a conservative effect size estimate of zero (e.g., Shadish, Montgomery, Wilson, Wilson, Bright, & Okwumabua, 1993).

Prior to analyses, the distributions of ESs and total sample size (*N*) were examined for the presence of outliers (i.e., any ES or *N* greater than or equal to three standard deviations beyond the mean; Shadish, Navarro, Matt, & Phillips, 2000). Seven outcome outliers and 10 *N*s were identified and then *windsorized*; that is, these values were reset to a value equaling three *SD*s from the mean.

We used a weighted least squares approach by following guidelines developed by Hedges and Olkin (1985). Comparisons between treatment and control groups were calculated using the standardized effect size (*g*),

and these were then weighted to adjust for small sample sizes (Hedges & Olkin, 1985). Treatment effects were calculated separately for each outcome category. If studies collected data on multiple measures within the same category, such as prosocial reasoning and prosocial decision making, the effect sizes for these outcomes were averaged to create a single effect for civic engagement.

A 0.05 probability level was used to detect statistical significance and 95% confidence intervals (CIs) were calculated around group means. Cumming and Finch's (2005) procedure was used to detect significant differences between group means by examining the extent of any overlap in the two groups' mean confidence intervals. A random effects model was used in the analysis to increase the generality of the findings (Lipsey & Wilson, 2001).

Results

Descriptive Characteristics

The 62 reviewed programs involved 11,837 students and almost half (48%) of the studies appeared after 2000. Sixty-seven percent of reports were published journal articles, while the remaining 33% were unpublished conference papers, dissertations, or technical reports. The majority of SL programs served college undergraduates (68%), while 16% involved high school students, and few programs served elementary (5%), middle school (5%), or graduate (6%) students. Of the 37 studies that mentioned participants' race or ethnicity, the predominant ethnicity was Caucasian in 16 studies (26%), mixed non-Caucasian groups in eight studies (13%), Latino in three studies (5%), and African American in one study (2%). The remaining nine studies reported serving a combination of Caucasian and non-Caucasian populations. Of the 40 studies that reported participants' gender, 34 (85%) reported a larger percentage of females. The ten studies that reported information on students' socioeconomic status served almost equal numbers of low-, middle-, and upper-class students.

In terms of methodology, 31% of studies used randomized designs and 41 studies (66%) included pretests. Out of the 380 total outcomes included in the 62 studies, 68% were based on reliable measures and 45% were drawn from valid measures. The majority of outcomes were student self-reports (87%), while the remaining data were derived from school records (7%) or based on outside observers (6%).

Findings for Student Outcomes

Table 1 presents the mean ESs and 95% CIs for each outcome category at post. Service-learning programs yielded statistically significant effects in all five areas: attitudes toward self, attitudes toward school and learning, civic engagement, social skills, and academic achievement

(mean ESs ranged from 0.27 to 0.43). Applying Cumming and Finch's (2005) procedure, the ES for academic achievement was significantly higher than the ES for the other four outcomes, which did not differ significantly from each other. These findings supported our first hypothesis that SL programs would be associated with multiple positive effects.

Table 1

Mean Effects for Student Outcomes

Outcome area	N	Mean ES	95% CI
Overall effect	62	0.28*	0.21–0.34
Attitudes toward self	36	0.28*	0.18–0.38
Attitudes toward school and learning	12	0.28*	0.12–0.43
Civic engagement	28	0.27*	0.16–0.38
Social skills	28	0.30*	0.18–0.38
Academic achievement	17	0.43*	0.29–0.58

Note. N = total sample size; ES = effect size; CI = confidence interval. The sum in the N column does not total 62 because some studies assessed outcomes in more than one area.

*Denotes that the mean effect is significantly different from zero at the 0.05 level.

Inclusion of Recommended Practices

Nine studies (15%) reported following four recommended practices, ten studies (16%) reported three, nine studies (15%) reported two, 21 studies (33%) reported one, and 13 studies (21%) apparently contained no recommended practices. Among all the practices we coded, the most frequent was the use of reflection (74%), followed by linking to curriculum (38%), youth voice (33%), and community involvement (26%). In 20 of the 21 studies that apparently used only one recommended practice, that practice was reflection.

Use of Recommended Practices and Outcomes

Although we hoped to analyze each outcome area separately in relation to the use of recommended practices, there were too few studies with which to do this, so outcomes from all categories were averaged within each study to yield one overall ES per study.⁴ Table 2 presents the mean

⁴ The fact that outcomes did not differ significantly across outcome domains suggests that using one average ES per study is acceptable.

effects and CIs for studies including zero, one, two, three, or all four recommended practices. Studies that reported using all four recommended practices yielded a significant positive effect ($ES = 0.35$) as did studies using one, two, or three recommended practices ($ESs = 0.30, 0.27,$ and $0.33,$ respectively). The 13 studies not using any of these practices also yielded a positive, but smaller, significant effect ($ES = 0.17$). But analyses indicate that the effects for studies using one, three, and four recommended practices were significantly higher than those studies using no recommended practices. Mean ESs of studies incorporating two practices were not significantly different from studies that used no recommended practices. In general, these findings supported our second hypothesis that studies including recommended practices would achieve significantly higher mean effects than those not following recommended practices. However, including more of the four recommended practices did not lead to successively better outcomes.

Table 2***Use of Recommended Practices and Student Outcomes***

Number of recommended practices	<i>N</i>	Mean ES	95% CI
None	13	0.17*	0.03–0.30
One	21	0.30*	0.19–0.41
Two	9	0.27*	0.12–0.43
Three	10	0.33*	0.16–0.50
Four	9	0.35*	0.18–0.52

Note. *N* = total sample size; ES = effect size; CI = confidence interval.

*Denotes that the mean effect is significantly different from zero at the 0.05 level.

Analyses of Methodological Characteristics

To assess the influence of methodological characteristics, we conducted separate analyses with studies grouped according to whether or not they met each methodological criterion (e.g., randomized design or not, and so on). Because of the few studies involving elementary students, we collapsed the educational level of the students into two categories: (K–12 versus college and beyond). Table 3 presents the results of these analyses, which indicate no significant differences between groups of studies on any of these variables. For example, the mean ESs of studies

with randomized designs and quasi-experimental design were virtually identical ($ESs = 0.31$ and 0.30 , respectively), and outcomes were comparable for students in the K–12 grades and in college or beyond. These additional analyses suggested that current findings were not being positively biased by less methodologically rigorous study procedures and that outcomes were comparable for students at all educational levels.

Table 3

Effects for Potential Moderators

Variable	<i>N</i>	Mean ES	CI
Randomization			
Yes	19	0.31*	0.18–0.43
No	43	0.30*	0.22–0.38
Pre/post testing			
Yes	41	0.29*	0.21–0.37
No	21	0.26*	0.15–0.36
Use of reliable outcome measures [†]			
Yes	260	0.23*	0.19–0.27
No	120	0.41*	0.35–0.47
Use of valid outcome measures [†]			
Yes	169	0.27*	0.22–0.32
Did not report	211	0.30*	0.26–0.35
Source of report [†]			
Self	330	0.28*	0.24–0.31
Other (observer, school record)	50	0.37*	0.28–0.47
Students' educational level			
K–12	19	0.20*	0.08–0.31
College	43	0.31*	0.23–0.39

Note. *N* = total sample size; ES = effect size; CI = confidence interval.

*Denotes that the mean effect is significantly different from zero at the 0.05 level.

[†]Each study could have several outcomes, so the *N* here is 380.

Discussion

As predicted, data from 62 studies indicate that, in comparison to controls, students participating in SL programs demonstrate significant gains in five outcome areas: attitudes toward self, attitudes toward school

and learning, civic engagement, social skills, and academic performance. These findings bolster the views of educators who posit that SL programs can benefit students at different educational levels in several ways. These multiple benefits include such areas as enhanced self-efficacy and self-esteem, more positive attitudes toward school and education, an increase in positive attitudes and behaviors related to community involvement, and gains in social skills relating to leadership and empathy. The relatively high mean effect for academic performance ($ES = 0.43^5$) is probably the most important finding for educators and advocates of SL programs. For example, the current political and administrative context of No Child Left Behind legislation puts pressure on schools to improve K–12 students' academic proficiency. The wider use of well-conducted SL programs could be one way to move toward this goal.

Also, as predicted, there was empirical support for the recent *K–12 Service-Learning Standards for Quality Practice* list, which emphasizes what elements should be included to improve the quality of SL programs. At least this was true for the four elements that we were able to examine: linking to curriculum, voice, community involvement, and reflection (National Youth Leadership Council, 2011). All studies, regardless of how many of the four recommended practices they contained, produced significant positive mean effects on the five outcomes (mean ESs ranging from 0.27 to 0.43). Moreover, programs that used all four practices yielded an overall mean ES that was twice the magnitude of programs using none of the four (0.35 versus 0.17, respectively). The results were not as completely straightforward as hoped, however, because using more of the four practices did not result in successively higher mean effects. That is, programs containing one practice seemed to be as successful overall (mean $ES = 0.30$) as those that contained two, three, or four ($ESs = 0.27, 0.33, \text{ and } 0.35$, respectively). In addition, the mean effects for programs containing two practices did not differ significantly from those containing none.

The findings suggest not only that the inclusion of some recommended practices is associated with more benefits for participants, but also that, in future research, there is a need to assess if some practices may be more important than others, and how the presence of multiple practices interacts with participant and other program characteristics to influence different outcomes. Moreover, reflection was the only recommended practice to be included in at least half of the studies, which suggests that current SL programs might be overlooking the potential importance of many recommended elements.

⁵ Cohen (1988) gives the following guidelines for the social sciences: Small effect size is typically between 0.1 and 0.23; medium is between 0.24 and 0.36; and large is 0.37 or greater.

Limitations and Recommendations for Future Research

This review involved a careful search for published and unpublished reports, included only studies with control groups, and found that studies with less-preferred methodological features (e.g., non-randomized designs, or the use of measures with questionable psychometric properties) were not associated with inflated effect sizes. These circumstances increase confidence in the main finding that participants in SL programs can benefit in multiple ways. Nevertheless, our review has limitations that suggest how future studies can be improved. Six suggestions can be offered.

First, complete reporting of study procedures is essential. Many reviewed reports contained incomplete or missing information on many important variables. We could only analyze the possible contribution of four of the eight elements in the *K-12 Service-Learning Standards for Quality Practice* list because of absent information on the other standards (diversity, meaningful service opportunities, program duration and intensity, and progress monitoring). These practices may have been followed, but authors discussed them too infrequently, if at all, to permit any statistical analyses. Furthermore, we could not explore the influence of participant characteristics (e.g., race/ethnicity or gender) because of missing data in many studies. In addition, more data using multiple outcome areas are needed, especially for academic achievement. Only 17% of studies included such outcomes. Providing more complete information about (a) the possible use of multiple recommended practices and (b) participant characteristics and assessing changes in multiple outcomes will allow for more penetrating analyses of SL programs and their effects.

Second, there were only a small number of controlled outcome studies involving elementary, middle school, or graduate students. This limits the generalizability of our results primarily to SL programs serving high school and college populations and suggests that future research should evaluate more programs for younger and older students. It is possible that some recommended practices are more important for younger (e.g., developmentally appropriate service opportunities) or older (e.g., youth voice) students.

Third, several methodological features could be enhanced in future research. Authors should strive to use more psychometrically sound assessments and randomized designs. One measure created specifically for the SL field and tested for validity and reliability is the Civic Attitudes and Skills Questionnaire (Moely, Mercer, Ilustre, Miron, & McFarland, 2002). It is understandable that some studies may not be able to randomize students because they might be selecting an SL course to graduate or to fulfill certain academic requirements; however, the lack of randomization introduces potential selection bias. Students who self-select into SL

programs may differ in important ways from those who are not interested in these programs and these differences might influence outcomes. Therefore, it is important to compare the initial status of SL and non-SL groups through pretesting. In addition to the use of psychometrically sound pre- and post-assessments and more randomized designs, it is important to collect follow-up data so that the durability of the impact of SL experiences can be estimated.

Fourth, too many studies (87% of the outcomes in this review) have relied on student self-report data. A college student's reported intention to vote may be very different than his or her actual voting behavior in university, local, or national elections. Similarly, students' ratings of commitment to their community could be biased by social desirability. New approaches are being developed, such as the item count technique, to correct for the social desirability bias with self-report measures (Holbrook & Krosnick, 2010). Self-report data can be useful, but it is preferable that they be complemented by other information drawn from peers, teachers, parents, or independent observers. Similarly, although it is heartening to know that students report that their academic learning improved during their SL experiences, it is essential to also document such gains with more objective information.

Fifth, we had to make several judgments about whether National Youth Leadership Council (2011) standards were being followed in SL programs, not only because relevant information was limited in the reports as already noted, but also because the standards are not clearly operationalized. For example, the standard regarding meaningful service (which was not assessed in this review due to absent information) emphasizes, among other things, that SL experiences should engage participants in meaningful and personally relevant service activities, link to academic curriculum, and incorporate ongoing reflection activities that prompt thoughtful analysis about oneself and one's relationship to society (2011). It is essential that members of the SL field provide more concrete guidelines on what it takes to achieve different standards so it is clear which standards are being met in each particular situation. Although the current standards are logically compelling, without greater clarity and specificity, it will not be possible to confirm if their inclusion directly leads to better program outcomes.

Sixth, and finally, investigations that attempt to identify what mediates changes in students would be extremely helpful. A recent study by Reinders and Yourniss (2009) is a good example. Their longitudinal study examined elements of adolescents' activities and how they experienced or interpreted these activities. Results supported their path-analytic model, which suggests that, over time, having direct interactions with people in need influenced adolescents' feelings of being helpful to others, which,

in turn, led to enhanced civic engagement. Additional studies that examine what leads to what during SL experiences would help others develop more effective programs.

In sum, this review provides evidence that SL programs have positive effects on students' attitudes, social behavior, and academic performance. Furthermore, the use of some recommended practices, such as reflection, seems to be associated with better outcomes. The findings should be gratifying to educators who incorporate SL into their courses, and should encourage further research to more fully understand the conditions that foster student growth and development in SL programs.

References

- Ammon, M. S., Furco, A., Chi, B., & Middaugh, E. (2002). *A profile of California's CalServe service-learning partnerships, 1997–2000*. Sacramento: California Department of Education.
- Anderson, S. (1998, September). *Service learning: A national strategy for youth development*. Washington, DC: George Washington University. Retrieved from www.gwu.edu/~ccps/pop_svc.html
- Astin, A. W., Vogelgesang, L. J., Misa, K., Anderson, J., Denson, N., Jayakumar, U., Saenz, V., & Yamamura, E. (2006). *Understanding the effects of service-learning: A study of students and faculty*. Retrieved from University of California, Los Angeles, Higher Education Research Institute website: http://www.heri.ucla.edu/PDFs/pubs/reports/UnderstandingTheEffectsOfServiceLearning_FinalReport.pdf
- Billig, S. H. (2002). Adoption, implementation, and sustainability of K–12 service-learning. In A. Furco & S. H. Billig (Eds.), *Service-learning: The essence of the pedagogy*. Greenwich, CT: Information Age.
- Billig, S. H. (2009). Does quality really matter: Testing the new K–12 service-learning standards for quality practice. In B. E. Moely, S. H. Billig, & B. A. Holland (Eds.), *Advances in service-learning research: Vol. 9. Creating our identities in service-learning and community engagement* (pp. 131–158). Greenwich, CT: Information Age.
- Billig, S. H., Root, S., & Jesse, D. (2005). The relationship between quality indicators of service-learning and student outcomes: Testing the professional wisdom. In S. Root, J. Callahan, & S. H. Billig (Eds.), *Advances in service-learning research: Vol. 5. Improving service-learning practice: Research on models that enhance impacts* (pp. 97–115). Greenwich, CT: Information Age.
- Blyth, D. A., Saito, R., & Berkas, T. (1997). A quantitative study of the impact of service-learning programs. In A. S. Waterman (Ed.), *Service learning: Applications from research* (pp. 39–56). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Boss, J. A. (1994). The effect of community service work on the moral development of college ethics students. *Journal of Moral Education*, 23, 183–98. doi:10.1080/0305724940230206
- Campus Compact. (2009). *2008 service statistics: Highlights and trends of Campus Compact's annual membership survey*. Boston, MA: Campus Compact.

Retrieved from <http://www.compact.org/wp-content/uploads/2009/10/2008-statistics1.pdf>

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conrad, D., & Hedin, D. (1982). The impact of experiential education on adolescent development. *Child and Youth Services, 4*, 57–76. doi:10.1300/J024v04n03_08
- Conway, J. M., Amel, E. L., & Gerwien, D. P. (2009). Teaching and learning in the social context: A meta-analysis of service learning's effects on academic, personal, social, and citizenship outcomes. *Teaching of Psychology, 36*, 233–245. doi:10.1080/00986280903172969
- Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60*, 170–180. doi:10.1037/0003-066X.60.2.170
- Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development, 82*, 405–432.
- Eyler, J., Giles, D. E., & Schmiede, A. (1996). *A practitioner's guide to reflection in service-learning: Student voices and reflections*. Nashville, TN: Vanderbilt University.
- Eyler, J., Giles, D. E., Stenson, C. M., & Gray, C. J. (2001). *At-a-glance: What we know about the effects of service-learning on students, faculty, institutions, and communities, 1993–2001*. Retrieved June 15, 2009, from the National Service-Learning Clearinghouse website: <http://www.servicelearning.org/pubs/materials/L054>
- Giles, D. E., & Eyler, J. (1994). The impact of a college community service laboratory on students' personal, social, and cognitive outcomes. *Journal of Adolescence, 17*, 327–339. doi:10.1006/jado.1994.1030
- Gray, M. J., Ondaatje, E. H., Fricker, R., Geschwind, S., Goldman, C. A., Kaganoff, T., & Klein, S. P. (1998). *Coupling service and learning in higher education: The final report of the evaluation of the Learn and Serve America, higher education program*. Santa Monica, CA: RAND Corporation.
- Hamilton, S. F., & Zeldin, S. R. (1987). Learning civics in the community. *Curriculum Inquiry, 17*, 407–420. Retrieved from <http://www.jstor.org/action/showPublication?journalCode=currinqu>
- Harwood, A. M., & Radoff, S. A. (2009). Reciprocal benefits of mentoring: Results of a middle school–university collaboration. In B. E. Moely, S. H. Billig, & B. A. Holland (Eds.), *Advances in service-learning research: Vol 9. Creating our identities in service-learning and community engagement* (pp. 131–158). Greenwich, CT: Information Age.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Boston, MA: Academic Press.
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly, 4*, 37–67. doi:10.1093/poq/nfp065
- Kramer, M. (2000). *Make it last forever: The institutionalization of service learning in America*. Washington, DC: Corporation for National Service.

- Larkin, R., & Mahoney, A. (2006). Empowering youth to change their world: Identifying key components of a community service program to promote positive development. *Journal of School Psychology, 44*, 513–531.
- Learn and Serve America. (2010). *What is service-learning?* Retrieved March 16, 2010, from <http://www.servicelearning.org/what-service-learning>
- Leming, J. S. (2001). Integrating a structured ethical reflection curriculum into high school community service experiences: Impact on students' sociomoral development. *Adolescence, 36*, 33–45. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=rzh&jid=ADO&site=ehost-live>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.
- Markus, G. B., Howard, J. P., & King, D. C. (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. *Educational Evaluation and Policy Analysis, 15*, 410–419. doi:10.3102/01623737015004410
- Melchior, A., & Bailis, L. N. (2002). Impact of service-learning on civic attitudes and behaviors of middle and high school youth: Findings from three national evaluations. In A. Furco & S. H. Billig (Eds.), *Advances in service-learning research: Vol. 1. Service-learning: The essence of the pedagogy* (pp. 201–222). Greenwich, CT: Information Age.
- Meyer, S. (2006). *Texas center for service-learning: Evaluation of K-12 SCP and CHESP programs*. Denver, CO: RMC Research.
- Moely, B. E., Mercer, S. H., Ilustre, V., Miron, D., & McFarland, M. (2002). Psychometric properties and correlates of the Civic Attitudes and Skills Questionnaire (CASQ): A measure of students' attitudes related to service-learning. *Michigan Journal of Community Service-learning, 8*(2), 15–26.
- Morgan, W., & Streb, M. (2001, March). Building citizenship: How student voice in service-learning develops civic values. *Social Science Quarterly, 82*, 155–169. doi:10.1111/0038-4941.00014
- National Research Council. (1999). *How people learn*. Washington, DC: National Academies Press.
- National Service-Learning Cooperative. (1998). *Essential elements of service learning*. St. Paul, MN: National Youth Leadership Council.
- National Youth Leadership Council. (2011). *K-12 service-learning standards for quality practice*. Retrieved from <http://www.nylc.org/k-12-service-learning-standards-quality-practice>
- Parker-Gwin, R., & Mabry, J. B. (1998). Service learning as pedagogy and civic education: Comparing outcomes for three models. *Teaching Sociology, 26*(4), 276–291.
- Perry, J. L., & Katula, M. C. (2001, July). Does service affect citizenship? *Administration and Society, 33*(3), 330–365.
- Reinders, H., & Yourniss, J. (2009). School-based required community service and civic development in adolescents. *Applied Developmental Science, 10*, 2–12. doi:10.1207/s1532480xads1001_1
- Shadish, W. R., Montgomery, L. M., Wilson, P., Wilson, M. R., Bright, I., & Okwumabua, T. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology, 61*, 992–1002. doi:10.1037/0022-006X.61.6.992

- Shadish, W. R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin*, *126*, 512–529. doi:10.1037/0033-2909.126.4.512
- Skinner, R., & Chapman, C. (1999). *Service-learning and community service in K–12 public schools*. Retrieved from the National Center for Education Statistics website: <http://nces.ed.gov/surveys/frss/publications/1999043/index.asp>
- White, A. (2001). *Meta-analysis of service-learning research in middle and high schools* (Unpublished doctoral dissertation). University of North Texas, Denton.

Copyright of Journal of Experiential Education is the property of Association for Experiential Education and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.