
James M. Conway
Central Connecticut State University

Elise L. Amel
University of St. Thomas

Daniel P. Gerwien
Central Connecticut State University

Service learning places teaching and learning in a social context, facilitating socially responsive knowledge. The purposes of this meta-analysis were to summarize evidence on (a) extent and types of change in participants in service learning programs, (b) specific program elements (moderators) that affect the amount of change in participants, and (c) generalizability of results across educational levels and curricular versus noncurricular service. We included 103 samples and found positive changes for all types of outcomes. Changes were moderate for academic outcomes, small for personal outcomes and citizenship outcomes, and in between for social outcomes. Programs with structured reflection showed larger changes and effects generalized across educational levels. We call for psychologists to increase their use of service learning, and we discuss resources for doing so.

Socially responsive knowledge involves experience-based education on social issues, including learning the skills to solve social problems. Authors such as Altman (1996) and Bringle and Duffy (1998) have proposed service learning as a way to instill socially responsive knowledge and increase civic engagement by linking the psychology curriculum to community needs. Service learning, according to Bringle and Hatcher (1995), is an educational experience involving an organized service activity with structured reflection to guide students’ learning. This focus embeds teaching and learning in a social context larger than a classroom. For example, Kretchmar’s (2001) general psychology students mentored at-risk school children and tutored boys in a transitional living facility. Structured reflection activities included small and large group discussions around issues such as the youths’ psychosocial development. Other types of reflective activities can include structured journal entries and papers integrating service experiences with course material (Lundy, 2007).

A strong theoretical basis for expecting service learning to benefit participants was provided by Kolb (1984) and Yates and Youniss (1996)—service learning should work because it supports the construction of knowledge through student reflection on experience, development of new conceptualizations, and experimenting with the new conceptualizations. Empirical reviews have drawn positive conclusions (e.g., Billig,
Table 1. Meta-Analysis Results by Type of Outcome

<table>
<thead>
<tr>
<th>Category</th>
<th>No. Studies</th>
<th>Total N</th>
<th>Mean d</th>
<th>95% Confidence Interval</th>
<th>True Standard Deviation of d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Academic outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Knowledge/GPA/grades</td>
<td>9</td>
<td>777</td>
<td>.42</td>
<td>.14–.71</td>
<td>.39</td>
</tr>
<tr>
<td>B. Cognitive outcomes</td>
<td>5</td>
<td>473</td>
<td>.29</td>
<td>.06–.52</td>
<td>.22</td>
</tr>
<tr>
<td>C. Academic motivation and attitudes</td>
<td>6</td>
<td>288</td>
<td>.58</td>
<td>-.10–1.26</td>
<td>.83</td>
</tr>
<tr>
<td>Personal outcomes</td>
<td>58</td>
<td>6,103</td>
<td>.21</td>
<td>.15–.27</td>
<td>.19</td>
</tr>
<tr>
<td>A. Self-evaluations</td>
<td>32</td>
<td>1,819</td>
<td>.26</td>
<td>.16–.37</td>
<td>.25</td>
</tr>
<tr>
<td>B. Volunteer motivations</td>
<td>7</td>
<td>555</td>
<td>.16</td>
<td>.00–.31</td>
<td>.16</td>
</tr>
<tr>
<td>C. Moral development</td>
<td>4</td>
<td>93</td>
<td>.34</td>
<td>.23–.44</td>
<td>.00</td>
</tr>
<tr>
<td>D. Alienation/deviance</td>
<td>13</td>
<td>3,070</td>
<td>.22</td>
<td>.12–.32</td>
<td>.14</td>
</tr>
<tr>
<td>E. Well-being</td>
<td>6</td>
<td>274</td>
<td>.17</td>
<td>-.07–.42</td>
<td>.26</td>
</tr>
<tr>
<td>F. Career development</td>
<td>15</td>
<td>1,890</td>
<td>.18</td>
<td>-.02–.39</td>
<td>.37</td>
</tr>
<tr>
<td>Social outcomes</td>
<td>37</td>
<td>3,271</td>
<td>.28</td>
<td>.18–.39</td>
<td>.27</td>
</tr>
<tr>
<td>A. Skill – interacting or working with others</td>
<td>15</td>
<td>2,370</td>
<td>.05</td>
<td>-.10–.20</td>
<td>.26</td>
</tr>
<tr>
<td>B. Understanding or tolerating diversity</td>
<td>17</td>
<td>2,097</td>
<td>.22</td>
<td>.04–.40</td>
<td>.34</td>
</tr>
<tr>
<td>C. Beliefs, knowledge, or attitudes toward those served</td>
<td>15</td>
<td>656</td>
<td>.44</td>
<td>.28–.60</td>
<td>.24</td>
</tr>
<tr>
<td>D. Beliefs or attitudes toward marginalized people in general</td>
<td>7</td>
<td>735</td>
<td>.13</td>
<td>.01–.26</td>
<td>.11</td>
</tr>
<tr>
<td>Citizenship outcomes</td>
<td>55</td>
<td>7,384</td>
<td>.17</td>
<td>.12–.23</td>
<td>.16</td>
</tr>
<tr>
<td>A. Personally responsible citizenship</td>
<td>12</td>
<td>1,608</td>
<td>.08</td>
<td>-.03–.20</td>
<td>.16</td>
</tr>
<tr>
<td>B. Participatory citizenship</td>
<td>18</td>
<td>2,070</td>
<td>.20</td>
<td>.12–.28</td>
<td>.11</td>
</tr>
<tr>
<td>C. Justice-oriented citizenship</td>
<td>17</td>
<td>1,662</td>
<td>.22</td>
<td>.03–.40</td>
<td>.37</td>
</tr>
<tr>
<td>D. Combination of citizenship types</td>
<td>47</td>
<td>6,526</td>
<td>.15</td>
<td>.09–.21</td>
<td>.16</td>
</tr>
</tbody>
</table>

2000, 2002; Eyler, Giles, Stenson, & Gray, 2001), but some authors have noted that the evidence regarding effects on participants is mixed (Eyler, 2002; Reinders & Youniss, 2006). The purposes of this study were to meta-analyze the existing evidence on (a) extent and types of change in participants in service learning programs; (b) specific program elements (moderators) that affect the amount of change in participants; and (c) generalizability of results across educational levels and curricular versus noncurricular service.

Types of Outcomes

We organized past research using four outcome categories based on several sources: Billig (2000, 2002), Eyler et al. (2001), and Eyler and Giles (1999). Major categories include (a) academic outcomes, (b) personal outcomes, (c) social outcomes, and (d) citizenship outcomes (see Table 1 for subcategories).

Academic Outcomes

Academic outcomes include cognitive and academic changes involving knowledge, ability to apply knowledge, cognitive processes, and motivation to learn. Billig (2002) and Eyler et al. (2001) reviewed supportive evidence on academic outcomes and noted that evidence on effects on grade point average (GPA) is mixed.

Personal Outcomes

Personal outcomes deal primarily with participants’ thoughts and feelings about themselves or their motives or values, and their well-being. Examples include self-esteem, self-efficacy, and career development. There is evidence that personal outcomes change as a result of service (Billig, 2002; Eyler et al., 2001; Yates & Youniss, 1996). However, some individual studies reported finding no evidence of a change (e.g., Johnson & Notah, 1999).

Social Outcomes

Social outcomes deal with participants’ relationships to others including skill in interacting with others (e.g., leadership skills) and thoughts and beliefs about others (e.g., attitudes toward the population one is serving). Evidence that social outcomes increase due to service experience is provided by Billig (2002), Eyler et al. (2001), and Yates and Youniss (1996). Yet, again some
studies have found no evidence of positive change (e.g., Moely, McFarland, Miron, Mercer, & Ilustre, 2002).

Citizenship Outcomes

Altman’s (1996) socially responsive knowledge implies a sense of citizenship, and Bringle and Duffy (1998) were explicit in pointing to democratic participation as a goal of psychology education. Our citizenship outcomes included three types based on Westheimer and Kahne’s (2004) descriptions: personally responsible (acting responsibly; e.g., obeying laws, recycling), participatory (active involvement in community improvement), and justice-oriented citizenship (addressing societal structures and injustice). Within each category, outcomes could involve actual behavior (e.g., frequency of volunteering), beliefs (e.g., about the importance of volunteering), or commitment or intentions to engage in the behavior. Several reviews provide evidence supporting changes in citizenship (Billig, 2002; Perry & Katula, 2001; Yates & Youniss, 1996). However, Perry and Katula (2001) mentioned that there have been studies showing no evidence of an effect, and Reinders and Youniss (2006) indicated that results are mixed.

Moderators of the Effect of Service Learning

Mixed results from previous research might be explained by differences in service learning implementation (Eyler, 2002). Therefore we examined which program elements contribute to greater change. We identified potential moderators cited in the literature—reflection (e.g., Eyler & Giles, 1999), direct contact with beneficiaries of service (Mabry, 1998), and intensity and duration of service (e.g., Tannenbaum & Berrett, 2005). Because almost all studies in our meta-analysis involved direct contact, we were not able to test it as a moderator and do not discuss it further.

Reflection

Reflection on one’s service experience has been described as “the hyphen in service-learning; it is the link that ties student experience in the community to academic learning” (Eyler & Giles, 1999, p. 171). Bringle and Hatcher (1999) described Dewey’s (1933) philosophical foundation for reflection in the learning process, and authors such as Eyler, Root, and Giles (1998), Marchel (2004), and Plante (1998) have described how reflection can be used to achieve academic learning goals (e.g., gaining content knowledge and problem-solving skills). We hypothesized that service will produce larger effects on participants when structured reflection is built in.

Intensity and Duration of Service

Dewey (1933), in discussing necessary conditions for experience to be educative, stated that the experience “must involve a considerable time span” (p. 218). This is consistent with the time-intensive learning process described by Kolb (1984). Eyler et al. (2001) and Tannenbaum and Berrett (2005) listed several sources as demonstrating the benefits of greater intensity and duration (e.g., Mabry, 1998). We therefore hypothesized that larger changes in participant outcomes will be associated with (a) greater number of hours and (b) greater length of service.

Purposes and Hypotheses

Our first purpose was to meta-analyze evidence of change in participants in service learning programs on four outcome types.

H1: We hypothesized that we would find evidence of change in participants on academic, personal, social, and citizenship outcomes.

Our second purpose was to examine specific program elements (moderators) that affect the amount of change in participants on all four types of outcomes.

H2: We hypothesized greater change for programs, including structured reflection, than for those not including reflection.

H3: We hypothesized greater change for programs of longer duration.

H4: We hypothesized greater change for programs with a greater number of service hours.

Our third purpose was to examine generalizability of effects across education levels (K–12 vs. higher education vs. adult and mixed groups) and curricular service (linked to an academic curriculum) versus noncurricular service; we did not have specific hypotheses and merely conducted exploratory analyses.
Method

Literature Search


We searched electronic databases, including Academic Source Premiere, Business Source Premiere, CINAHL and Pre-CINAHL, ERIC, Dissertation Abstracts, Hospitality and Tourism Index, MEDLINE, PAIS International, Professional Development Collection (education), PsycINFO, and SocINDEX using the following keywords: (volunteerism or community service or service learning) and (effect* or outcome* or impact) and (longitudinal or pretest or posttest or change or increase) NOT community services. We also conducted a search of the library catalog on the National Service Learning Clearinghouse Web site (www.servicelearning.org).


We contacted researchers and service learning professionals through e-mail, and submitted requests to three electronic mailing lists available through the National Service Learning Clearinghouse Web site (www.servicelearning.org).

Criteria for Inclusion

To be included, a study had to meet the following criteria: (a) pretest–posttest design using identical quantitative measures for identical pre- and postsamples; (b) participation in community service between pre- and posttests; (c) sufficient information provided so we could classify the measures (e.g., as a citizenship or social outcome); and (d) reporting of pretest and posttest means, the pretest standard deviation, and sample size. In some cases, we contacted the authors to obtain information. Our final data set included 103 independent samples reported in 78 separate sources identified in the reference list by asterisks.

Coding the Studies

Two of the authors independently coded each study using detailed coding instructions, and for any disagreements we reached consensus through discussion.

Statistical information. Pretest and posttest means and pretest standard deviation were recorded for each outcome variable so that we could compute a d statistic. This information was straightforward and only required one coder. In some cases we were able to calculate d from other information (e.g., a t value).

Education level. We recorded the population of participants as: (a) higher education (college or university) students (46 studies), (b) K–12 students (42 studies, 3 from elementary school, 6 from middle school, and 33 from high school), or (c) adults, a mix of populations, or other (15 studies). Coders agreed on 98% of studies.

Curricular vs. noncurricular service. We classified each study according to whether the service was (a) part of a course (curricular service; 77 studies) or (b) not part of a course (noncurricular service; 26 studies). Coders agreed on 88% of studies. Of the 77 curricular service studies, 24 dealt with psychology or a related discipline.

Reflection. We recorded whether the authors (a) did not mention structured reflection techniques (38 studies), or (b) indicated that at least one structured reflection technique was used (65 studies). Coders agreed on 92% of studies.

Number of hours. We categorized each study as having (a) 40 or fewer hr of service (53 studies) or (b) 41 or more hr (22 studies). The 40-hr cutoff point represents one full-time work week and gave us a reasonable split in terms of number of studies in each category. Twenty-eight studies could not be coded due to lack of information. Coders agreed on 80% of studies. We also conducted more fine-grained analyses using four categories: (a) 10 or fewer hr (11 studies), (b) 11 to 15 hr (14 studies), (c) 16 to 40 hr (18 studies), and (d) 41+ hr (23 studies).

Number of weeks. We categorized each study as having (a) 15 or fewer weeks of service (59 studies) or
The 15-week cutoff represents one college semester and gave us fairly large numbers of studies in each category. Ten studies could not be coded because they did not provide the information. Coders agreed on 81% of studies. We conducted more fine-grained analyses using four categories: (a) less than 10 weeks (19 studies), (b) 10 to 15 weeks (44 studies), (c) 16 to 30 weeks (8 studies), and (d) 31+ weeks (20 studies).

**Outcome variables.** To classify outcome variables (e.g., as personal outcomes, social outcomes) we used the set of categories presented in Table 1 (our coding categories included detailed descriptions distinguishing different categories). We coded a total of 415 separate outcome variables (most studies had multiple outcome variables). Of the 103 studies, 19 had an academic or learning outcome, 58 had a personal outcome, 37 had a social outcome, and 55 had a citizenship outcome. Coders initially agreed on 63% of the category codes; as mentioned previously, all disagreements were resolved through discussion.

**Effect Sizes**

We calculated Cohen's $d$ (Cohen, 1988) as the difference in means (posttest minus pretest) divided by the pretest standard deviation (Morris & DeShon, 2002). Many studies had two or more outcome variables receiving the same category code. In these cases, we computed separate $d$ values for each outcome variable and then averaged the $d$ values within the same category. In some cases (e.g., measures of alienation or deviance) we reversed the signs of $d$s so that higher $d$ values always indicated positive changes.

We did separate meta-analyses for each type of outcome using the Comprehensive Meta Analysis software (Borenstein, Hedges, Higgins, & Rothstein, 2007). Our main results were based on random-effects rather than fixed-effects meta-analysis as described by Hedges and Vevea (1998) and Morris and DeShon (2002). Using fixed-effects analysis would mean results could not be generalized to future service programs in different locations, or with different groups of students, whereas the random-effects approach allows us to generalize. In the random-effects approach, sampling error variance is estimated differently, yielding wider confidence intervals, as compared to fixed-effects analysis. Overton (1998) demonstrated that random-effects meta-analysis can overestimate confidence intervals and have low power for moderator analysis. Therefore, for moderator effects, we report significance test results for both random- and fixed-effects approaches (consistent with other meta-analyses, such as Bettencourt, Talley, Benjamin, & Valentine, 2006).

We calculated the mean effect size for each outcome category, with each $d$ value weighted by the inverse of its sampling error variance. We also calculated 95% confidence intervals and the estimated “true” standard deviation of effect sizes (removing variation due to random sampling error).

**Results**

**Types of Outcomes**

Hypothesis 1 stated that we would find evidence of change in participants on academic, personal, social, and citizenship outcomes. Mean effect sizes for each specific outcome category are shown in Table 1, along with number of studies, total sample size, 95% confidence interval, and true standard deviation of effect sizes. Readers should use caution when interpreting results for categories with small numbers of studies or small total sample sizes.

**Academic outcomes.** Overall academic and learning outcome results showed a moderate effect size with a mean weighted $d = .43$, almost half a standard deviation. The confidence interval did not include zero, analogous to finding that the mean $d$ is statistically significantly greater than zero. The true standard deviation of .44 indicates large variation across programs in effect size. For subcategories, academic motivation, attitudes, and knowledge, and GPA, and grades had considerably higher means (.58 and .42, respectively) than did cognitive outcomes ($M = .29$). However, small numbers of studies and wide confidence intervals make it imprudent to draw firm conclusions about these differences.

**Personal outcomes.** The overall results for personal outcomes showed a small effect with a mean weighted $d = .21; the confidence interval did not include zero. The true standard deviation of .19 indicated that there was variation in effect sizes across programs. Subcategories for personal outcomes had mean weighted effect sizes ranging from .16 (volunteer motivations) to .34 (moral development).

**Social outcomes.** Social outcomes showed a fairly small effect with a weighted mean $d$ of .28; the
confidence interval did not include zero. Subcategory results in Table 1 show the lowest effect size for social skills (e.g., interpersonal and leadership skills) with mean $d = .05$, and the highest for beliefs, knowledge, or attitudes toward those served with mean $d = .44$; true standard deviations indicated variation across programs.

Citizenship outcomes. Citizenship outcomes showed the smallest effect with a mean $d$ of .17 (significantly different from zero). The subcategories did not show any effect sizes higher than .22, and true standard deviations indicated variation across programs.

Comparisons among outcome categories. We did not hypothesize about which outcome categories would show the largest changes but conducted exploratory comparisons using the $Q$ statistic which is distributed like $\chi^2$. The largest mean $d$ for any of the four categories was for academic or learning outcomes ($d = .43$), which was significantly larger than the effect for personal outcomes ($d = .21$), $Q(1) = 3.53$, $p = .06$ for random effects and $Q(1) = 14.98$, $p < .05$ for fixed effects. It was also larger than the effect for citizenship outcomes ($d = .17$), $Q(1) = 5.04$ and 18.75, both $p < .05$, for random and fixed effects, respectively. Academic or learning and social outcomes were not significantly different. Social outcomes ($d = .28$) had the second largest effect size, which was significantly larger than that for personal outcomes (only according to the fixed-effects analysis), $Q(1) = 6.36$, $p < .05$, and citizenship outcomes, $Q(1) = 3.43$, $p = .064$ for random effects and $Q(1) = 9.54$, $p < .05$ for fixed effects.

In summary, Hypothesis 1 (that we would find positive changes in all outcome categories) was supported, although changes were largest for academic and social outcomes, and smaller for personal outcomes and citizenship outcomes.

Moderator Variables

We performed moderator analyses only for the overall category results, and we excluded academic or learning outcomes because it was not prudent to divide further the 19 studies.

Reflection. Results were generally consistent with Hypothesis 2, stating that structured reflection would produce greater changes in outcomes (see Table 2). Differences between programs with and without structured reflection were relatively large for personal outcomes, $Md = .29$ with reflection and $Md = .09$ without; both $d$ values had confidence intervals that did not include zero ($Q$ values were 11.142 and 17.916 for random and fixed effects, both $ps < .05$).

For social outcomes the difference in mean $d$ values of .37 with reflection and .17 without was large (and both confidence intervals did not include zero), but only the random-effects test approached significance, $Q = 3.348$, $p = .067$. Citizenship outcomes had a mean of .22 with reflection versus .12 without reflection (both confidence intervals did not include zero); the fixed-effects test showed a significant difference, $Q = 7.285$, $p < .05$, and the random-effects test approached significance, $Q = 3.237$, $p = .072$.

Intensity and duration of service. As stated in Hypotheses 3 and 4, we expected larger effects with larger numbers of hours and weeks. Mean $d$ values for number of service hours (40 or less vs. 41 or more) and number of weeks of service (15 or less vs. 16 or more) did not support our hypotheses; $d$ values were slightly larger for smaller numbers of hours and weeks. We probed results using narrower categories and trends suggested increasing effects up to about 40 hr and 30 weeks, with lower effects for higher numbers of hours and weeks. However, the trends were not crystal clear and the categories often had very small numbers of studies. We also considered looking at combinations of hours and weeks, for example, to see if highly concentrated service (high hours in few weeks) is different from more extended service. We were unable to do this because very few studies showed highly concentrated or diffuse patterns.

Generalizability of Effects

Education level. Table 3 provides comparisons between K–12 students, higher education students, and adult and mixed groups (only for overall category results). We had no hypothesis but compared the three groups separately for personal, social, and citizenship outcomes (we did not make comparisons for academic outcomes). For personal and social outcomes, K–12 and higher education groups showed significant effects in the .20s or .30s. Q tests (both random and fixed effects) showed these to be significantly larger than mean $d$ values for adult or mixed groups, which had mean effect sizes near zero. For citizenship, K–12 had a low mean $d$ of .09, although the confidence interval did not include zero, whereas higher education had a significantly larger (according to both random- and fixed-effects Q tests) mean $d$ of .30. Adult or mixed
Table 2. Meta-Analysis Results for Structured Reflection

<table>
<thead>
<tr>
<th>Outcome Category</th>
<th>No. Studies</th>
<th>Total N</th>
<th>Mean d</th>
<th>95% Confidence Interval</th>
<th>True Standard Deviation of d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reflection</td>
<td>21</td>
<td>3,113</td>
<td>.09$^a$</td>
<td>.01–.17</td>
<td>.12</td>
</tr>
<tr>
<td>Reflection</td>
<td>37</td>
<td>2,990</td>
<td>.29</td>
<td>.20–.38</td>
<td>.23</td>
</tr>
<tr>
<td>Social outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reflection</td>
<td>16</td>
<td>1,630</td>
<td>.17$^b$</td>
<td>.01–.33</td>
<td>.29</td>
</tr>
<tr>
<td>Reflection</td>
<td>21</td>
<td>1,641</td>
<td>.37</td>
<td>.23–.51</td>
<td>.27</td>
</tr>
<tr>
<td>Citizenship outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No reflection</td>
<td>23</td>
<td>4,546</td>
<td>.12$^a$</td>
<td>.05–.19</td>
<td>.14</td>
</tr>
<tr>
<td>Reflection</td>
<td>32</td>
<td>2,839</td>
<td>.22</td>
<td>.14–.31</td>
<td>.19</td>
</tr>
</tbody>
</table>

$^a$The mean $d$ values for no reflection versus reflection were significantly different according to both random- and fixed-effects $Q$ tests (for citizenship the random-effects $p$ value was .072). $^b$Differences in mean $d$ values were significantly different according to random-effects test only ($p = .067$).

...groups had a mean $d$ of .21 for citizenship; the fixed-effects but not random-effects test showed this value to be significantly different from both other groups. In summary, effects generalized across K–12 and higher education for personal and social outcomes but did not generalize to adult or mixed participants.

Curricular vs. noncurricular service. For personal outcomes curricular service (41 studies vs. 17 for noncurricular service) had a significantly higher mean $d$ value, .27 versus .06 (both fixed- and random-effects $Q$ tests were significant, $p < .05$). Social and citizenship outcomes did not show significant differences, although mean $d$ values were higher for curricular service (.33 vs. 12 for social and .18 vs. 15 for citizenship outcomes). Therefore there is some evidence that noncurricular service has smaller effects. It is important to note that almost all adult studies were of noncurricular community service, but comparisons excluding adults still showed the same curricular–noncurricular differences.

Another confounding factor, for both curricular service and adult studies, was reflection: Most noncurricular studies and most adult studies did not use structured reflection. We therefore replicated comparisons for curricular versus noncurricular service, and adult versus other programs, within each level of reflection. Mean differences were almost all still in the same direction we have reported but most were nonsignificant, although this might be due to small numbers of studies in some categories.

Discussion

Service learning places teaching and learning in the social context of the community, and has been

Table 3. Meta-Analysis Results for Education Level

<table>
<thead>
<tr>
<th>Outcome Category</th>
<th>No. Studies</th>
<th>Total N</th>
<th>Mean d</th>
<th>95% Confidence Interval</th>
<th>True Standard Deviation of d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Personal outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K–12</td>
<td>29</td>
<td>3,458</td>
<td>0.25$^a$</td>
<td>.18–.32</td>
<td>.16</td>
</tr>
<tr>
<td>Higher education</td>
<td>19</td>
<td>992</td>
<td>0.28$^a$</td>
<td>.13–.43</td>
<td>.29</td>
</tr>
<tr>
<td>Adult/mix/other</td>
<td>10</td>
<td>1,653</td>
<td>0.00$^b$</td>
<td>−.12–.13</td>
<td>.16</td>
</tr>
<tr>
<td>Social outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K–12</td>
<td>5</td>
<td>294</td>
<td>0.37$^a$</td>
<td>.17–.58</td>
<td>.19</td>
</tr>
<tr>
<td>Higher education</td>
<td>22</td>
<td>1,296</td>
<td>0.36$^a$</td>
<td>.22–.51</td>
<td>.29</td>
</tr>
<tr>
<td>Adult/mix/other</td>
<td>10</td>
<td>1,681</td>
<td>0.08$^b$</td>
<td>−.12–.28</td>
<td>.29</td>
</tr>
<tr>
<td>Citizenship outcomes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K–12</td>
<td>26</td>
<td>4,128</td>
<td>0.09$^a$</td>
<td>.02–.15</td>
<td>.13</td>
</tr>
<tr>
<td>Higher education</td>
<td>19</td>
<td>1,594</td>
<td>0.30$^a$</td>
<td>.19–.42</td>
<td>.20</td>
</tr>
<tr>
<td>Adult/mix/other</td>
<td>10</td>
<td>1,663</td>
<td>0.21$^a$</td>
<td>.09–.33</td>
<td>.14</td>
</tr>
</tbody>
</table>

Note. Mean $d$ values for each outcome that do not share the same subscript differ at $p < .05$ on both random-effects and fixed-effects $Q$ tests.

$^a$Means for adult/mix/other differ significantly from other means on fixed-effects but not random-effects $Q$ tests.
proposed as a way to promote socially responsive knowledge, self-efficacy and self-esteem, compassion, and political participation (Altman, 1996; Bringle & Duffy, 1998). Our first major finding is that service learning does tend to have these effects, producing positive changes in academic, personal, social, and citizenship outcomes. We found the largest changes for academic outcomes and for beliefs, knowledge, or attitudes toward those being served. Changes for personal and citizenship outcomes were small. A second major finding was that reflection was generally associated with larger effects, and a third finding was that effects tended to generalize across K–12 and higher education programs, although for adult or mixed populations there was little evidence of changes in personal or social outcomes. There was some evidence that noncurricular service had smaller effects.

One limitation of our meta-analysis is that most studies were from disciplines other than psychology (about 23% were from psychology or a related discipline). We conducted analyses for psychology and related disciplines versus other disciplines and found that the major results held for both categories. We are therefore confident that our findings can be applied to the teaching of psychology.

**Recommendations for Teachers of Psychology**

**Consider using service learning.** We realize this is easier said than done, as service learning requires substantial effort to plan and carry out. Luckily, the psychology literature includes excellent examples to emulate. For general psychology and life-span development students, Kretchmar (2001) and Lundy (2007) provided a variety of service options including mentoring at-risk youths, working with homeless families (Kretchmar, 2001), and working in preschool or assisted living environments (Lundy, 2007). Reflection included group discussion with structured questions (e.g., linking clients’ needs to Maslow’s hierarchy) for Kretchmar (2001), and journals, papers, and oral presentations for Lundy (2007). In an applied animal behavior course, Kogan and Kellaway (2004) closely integrated classroom discussion, summary and reaction papers, and a term paper with twice-weekly meetings at the local humane society where students applied learning principles to train dogs in preparation for adoption. In Stadtlander’s (2002) graduate cognition course, each student met with an older adult in an assisted living facility every 2 weeks to collect data for experiments and allow participants to exercise their cognitive skills (producing evidence of cognitive improvement over the semester). Other examples include courses in research methods (Chapdelaine & Chapman, 1999), psychology ethics (Plante, 1998), and pediatric psychology (Hardy & Schaein, 2000).

Teachers new to service learning will need guidance on its nuts and bolts, such as identifying placements, establishing student and community partner expectations, monitoring progress, designing reflective activities, and so on. An excellent resource for getting started is the online National Service Learning Clearinghouse (www.servicelearning.org). We also recommend talking to as many colleagues involved in service learning as possible about their experiences.

Many colleagues will likely mention time constraints as a barrier. This constraint is very real but can be reduced through creative solutions such as using internal clients (Heckert, in press). Another constraint is that some service-learning efforts require funds for materials, transportation, faculty development, and so on. There are many grants and resources available to fund faculty development and actual service work through organizations such as the National Service-Learning Partnership (www.service-learningpartnership.org), Youth Service America (www.ysa.org), and the aforementioned National Service Learning Clearinghouse.

**Consider which outcomes to target.** Any one of the major types of outcomes is viable for service learning courses, but we suggest targeting particular outcomes and designing service learning experiences appropriately. We believe academic or learning outcomes will virtually always be appropriate. We also believe that whenever students have direct contact with members of marginalized groups there is an excellent opportunity for changing stereotypical attitudes and beliefs. Our experience is that students naturally begin to think about the people they serve, and are generally open to reflective activities about their own stereotypes. Examples of activities include identifying students’ implicit stereotypes using the Implicit Association Test (available on the World Wide Web; Project Implicit, 2008) and having students think about the world from their clients’ perspectives using a stereotype reduction technique studied by Galinsky and Moskowitz (2000).

For achieving other goals (e.g., increasing citizenship or interpersonal skills), service and reflective activities will need to be carefully designed (Battistoni, 2001). For example, if a goal is to stimulate interest in the political process, appropriate service and reflection activities could focus on direct political engagement.
(e.g., working on local mayoral election campaigns, voter registration). For citizenship in particular, the very small effect of service for the K–12 group suggests the possibility that youths are not cognitively or socially mature enough to easily connect their experiences to their civic beliefs and intentions. An exploratory follow-up analysis showed a significantly larger $d$ value for high schoolers (19 studies; $d = .14$) than for elementary and middle schoolers (7 studies; $d = -.05$), so the concern might apply mainly to the youngest students.

**Use carefully designed reflection techniques.** Our results showed the value of structured reflection, and there are several good resources available to guide teachers. Specific techniques are described by Dunlap (1998) and Bringle and Hatcher (1999). Structured reflection can come in the form of journals in which students connect their experiences with specific course-related concepts or thought-provoking questions. Additional options include in-class discussion or debates that are focused on connecting experience with specific course goals, and written assignments and research papers. The effectiveness of these techniques can be further enhanced through frequent, specific feedback from classmates, faculty, and even those served (Higgins, Hartley, & Skelton, 2002). Other good resources for designing reflection to maximize learning include Eyler, Giles, and Schmiede (1996) and Rama and Battistoni (2001). Future research should investigate how reflection can best be structured to facilitate particular outcomes. This will help teachers of psychology to leverage the power of service learning.

**References**


Learning through doing: An assessment of the relation between service learning attitude and effective course connections. Paper presented at the annual meeting of the Association for Psychological Science, Chicago.


*Mabry, J. B. (1998). Pedagogical variations in service-learning and student outcomes: How time, contact, and...


* Indicates a study included in the meta-analysis.

Notes

1. We are grateful to Chris Boyatzis and JoAnn Campbell for their comments on an earlier version of this article.

2. Send correspondence to James M. Conway, Department of Psychology, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050–4010; e-mail: conwayj@ccsu.edu