Revisiting Pedagogical Variations in Service-Learning and Student Outcomes

Thomas A. Dahan

Rutgers University-Camden

This article presents results from a quasi-experimental replication study that extended Mabry's 1998 correlational study by introducing a counterfactual (using psychometrically defensible measures) and controlling for academic motivation and relationships among students, faculty, and staff. The study offered evidence of the effectiveness of frequency of contact, reflection, and service duration in influencing civic and academic outcomes. Though the study's quasi-experimental approach posed modest limitations, the measured effects were similar to those identified in Mabry's study, thus improving confidence around findings related to the impacts of service-learning pedagogical variations and their implications for instructional design.

Keywords: replication, service-learning, pedagogy, quasi-experiment

Service-learning studies have demonstrated that student outcomes are related to participation in quality programs (Eyler & Giles, 1997, 1999), learning environments (Bringle, Hatcher, & Muthiah, 2010; Moely & Ilustre, 2014), and pedagogical variations in service- learning (Mabry, 1998). Mabry (1998) used three independent variables—time spent engaging in service-learning activities, contact with service beneficiaries, and student reflection—to compare results among service-learners regarding their civic attitudes and personal social values. Mabry found positive relationships associated with frequent contact with beneficiaries, participation in 15 to 19 hours of service-learning activities, ongoing and summative written reflection, and weekly in-class reflection. However, Mabry's study suffered many limitations common to service-learning research, including small sample size, lack of a counterfactual, and use of measures that contained limited psychometric evidence (Bringle, Phillips, & Hudson, 2004; Steinberg, Bringle, & McGuire, 2012). This study both confirmed and expanded Mabry's research by testing a larger, more diverse sample, introducing a counterfactual group to the quasi-experimental design, and using additional psychometrically defensible measures as outcomes and covariates.

Replication in Service-Learning Research

Adams, Ajrouch, Henderson, and Heard (2005) studied replication of service-learning research. Replication, the authors suggested, "is an essential aspect of scientific production and necessary for establishing consistent social patterns, reducing experimenter bias and increasing objectivity, identifying normative patterns independent of deterministic explanations, and consideration of the role of 'context' with regards to empirical findings" (p. 58). These scholars reviewed 44 service-learning studies and identified only two whose explicit purpose was to replicate the results of other scholars' work. They categorized replications as "literal," "operational," "instrumental," and "constructive." Literal replications involve attempts to replicate both the measurement of a study's outcome and its treatment. Operational replications change the measure of the outcome variable(s) but attempt to replicate the treatment and the independent variable(s). Instrumental replications change the independent variable(s) but replicate the outcome variable(s). A constructive replication attempts to improve both the independent and dependent variables while maintaining the conceptual integrity of the original study.

Bringle (2003) noted that good theories are based on hypotheses that are falsifiable, drawing on Karl Popper's hypothetico-deductive tradition of social science (Delanty & Strydom, 2003). This study

attempted to retest the theory that time, contact, and reflection influence student outcomes as they relate to service-learning activities. The results of the study add to the knowledge base of service-learning by contributing a new dataset of similar as well as different outcome measures using similar independent variables to reproduce results from an earlier correlational study of service-learning. Consistent with the recommendations of Adams et al. (2005), this study qualified as operational replication because it replicated most of the independent measures from Mabry's (1998) original study. It was also constructive because it introduced new independent variables that helped correct for challenges to internal validity, introduced the counterfactual to the quasi-experimental design, and used more reliable dependent measures for academic outcomes than the original study.

Quasi-Experiments in Service-Learning Research

The use of in service-learning research is well documented (Bringle & Steinberg, 2010). Quasi-experimental designs are suitable for service-learning research because they use intact groups in cases in which random assignment is difficult to achieve (Creswell, 2008). Steinberg et al. (2012) highlighted the problem of self-selection bias in quasi-experiments, and, indeed, Mabry (1998) identified self-selection as a potential limitation of her own work. As she stated, "self-selection is not particularly problematic here because the goal of this study is not to compare outcomes of service-learners against those who do not participate in service-learning" (p. 34). However, by electing to not use a control group, she limited the robustness of her findings. The current study improved Mabry's design by introducing a counterfactual: a control group consisting of students enrolled in traditional courses at the same level and in the same discipline as the service-learning treatment. While the inclusion of the control group did not resolve the self-selection bias, it offered a better opportunity to isolate the true effects of participation in service-learning (Kenny, 1975; Steinberg et al., 2012).

Method

The nonequivalent control group design of this study used data collected over three semesters through a pen-and-paper survey administered at the beginning and end of each semester between the spring of 2013 and the spring of 2014. A total of 606 complete responses to the pretest and 427 complete responses to the posttest dependent variables were collected. However, due to non-response on certain independent variable measures, the final sample consisted of 170 service-learners and 167 non-service-learners, representing a final response rate of 55% but an increase in sample size compared with Mabry's (1998) original study (n = 144). The descriptive statistics for the final sample are presented in Table 1, along with information from Mabry's sample.

Table 1. Study Sample Characteristics Compared to Mabry's (1998) Sample

Characteristic	Mabry	Sample		
		Control	S-L	Total
Male	32	34.3	41.7	38.0
Female	68	65.7	58.3	62.0
African American	2	10.7	16.4	13.6
American Indian	0	0	0	0
Asian	8	7.2	6.5	6.8
Hispanic	3	7.2	10.0	8.6
Two or More	2	4.2	5.2	4.7

Unknown	1	4.9	0.7	0.8
No Previous College	n/a	38.9	40.0	39.5
Mean Age	20	23	23	23
N	144	167	170	337

Note. Percentages reported for gender, ethnicity, transfer (previous college) status.

The survey collected demographic information including age, gender, ethnicity, years of college enrollment, previous college(s) attended, and information about previous participation in service-learning. This sample was more diverse than Mabry's (1998) regarding age, gender, and ethnicity. Tests for equivalence indicated that the treatment and control samples in the present study were statistically similar in composition using bivariate linear regression, which is the statistical equivalent of two-sample t-tests.

The sample was drawn from 33 different courses, including classes in art, psychology, sociology, urban studies, English, childhood studies, business, and social work. The courses with service-learning components whose students entered the treatment group possessed great variability of activities related to service-learning. To that end, these courses were included in the study in order to maximize the observable impacts of that variability on student outcomes. The control group of courses was selected based on similarity of discipline and subject area. For example, a course on youth identities was paired with a course on child wellbeing; a course in ethics and socially responsible business was paired with a course on business civic engagement. In a few cases, courses actually comprised the same topic (e.g., versions of social welfare policy, experimental psychology, and organizational behavior with and without service-learning).

Dependent Variables

This study used three scales to measure student civic engagement outcomes: a five-item personal social values scale (Mabry, 1998), a five-item civic attitudes scale (Mabry, 1998), and five of the 10 items from the community service self-efficacy scale (Reeb, Katsuyama, Sammon, & Yodder, 1998). Bringle et al. (2004) commented that Mabry's (1998) scales show promise as measures of student civic engagement, but there was limited psychometric evidence available regarding the construct validity of these measures. The addition of the community service self-efficacy items provide additional reliability evidence to measure student civic engagement, given the extensive validation of the community service self-efficacy instrument (Reeb, 2006; Reeb et al., 1998; Reeb, Folger, Langsner, Ryan, & Crouse, 2010).

In addition, 13 items from the National Survey of Student Engagement (NSSE) (Indiana University, 2011) were used to measure academic gains in general education, practical competencies, and personal and social development. These scales were each selected to measure academic outcomes because they can be benchmarked to national studies of student learning and engagement. The reliability information and descriptive statistics (raw mean score) for the scales is presented in Table 2. The Cronbach alpha levels meet acceptable levels of social science research (Devellis, 2012). For the final analyses, dependent variables (both pretest and posttest) were standardized with a mean of 0 and standard deviation of 1 to compare magnitude of the effects on the same scale.

Table 2. Internal Consistency Reliability of Dependent Variables

Multi-Item Scale	*N*	Cronbach's α		*M* (SD)	
		Pre	Post	Pre	Post

Civic Engagement					
Personal Social Values [1]	311	.77	.79	3.14 (.57)	3.22 (.57)
Civic Attitudes [1]	335	.74	.82	4.07 (.53)[4]	4.15 (.58)
Community Service Self-Efficacy [2]	335	.90	.91	7.77 (1.68)[4]	8.07 (1.56)
Academic Gains					
General Education [3]	326	.82	.83	3.15 (.61)	3.24 (.65)
Practical Competencies [3]	326	.70	.71	3.05 (.61)[4]	3.16 (.60)
Personal and Social Development [3]	326	.73	.72	2.81 (.66)	2.99 (.62)

Note.
[1] Originally reported in Mabry (1998).
[2] Five items adapted from the scale reported in Reeb, Katsuyama, Sammon, & Yodder (1998).
[3] Adapted from scales used in the National Survey of Student Engagement (Indiana University, 2011).
[4] Service-learners had higher pretest scores in practical competencies, civic attitudes, and community service self-efficacy compared with non-service-learners. No detectable differences were present for social values or the other academic gains measures.

Independent Variables

The three-item relationships scale from NSSE (Indiana University, 2011) was used to measure student relationships with others. This scale had a Cronbach's alpha of .71 with 334 observations. An eight-item academic motivation scale (Pascarella, 2007) was also used, with a Cronbach's alpha of .81 with 333 observations. These two scale measures were included as covariates in analyses because academic motivation and relationships predictably influence academic outcomes and may also influence student civic engagement.

This study used several similar independent variables employed in Mabry's (1998) original study examining how time, contact, and frequency of reflection influence service-learning outcomes. However, due to space considerations on the pen-and-paper form, this study did not include Mabry's questions regarding the kinds of in-class and out-of-class reflection. For in-class and out- of-class reflection, indicator variables were produced for weekly reflection and less frequent reflection. For contact, indicators were created for frequent (most or all of service time) and less frequent contact with community members. Indicator variables were generated for service hours based on the quartile levels (15 or fewer, 16-24, 25-30, and greater than 30). For each of these independent variables, the reference case was the control group.

Analysis

The study utilized ordinary least squares regression to permit statistical controls necessary for interpreting nonrandom data. Robust standard errors were calculated to correct heteroscedasticity or the inequality of variance across levels of the dependent variable (Gujarati & Porter, 2009). To that end, a less conservative alpha level of 0.10 was selected to accommodate the wider standard errors that resulted.

Kenny (1975) suggested that quasi-experimental data may be analyzed using analysis of covariance (ANCOVA), ANCOVA with reliability correction, raw change score analysis, or standardized change

score analysis. Generally, due to limitations in the quasi-experimental design that constrain internal validity, the procedure for analysis must be selected to accurately model the treatment effects. Researchers should therefore compare models to ensure that the effects are correctly modeled.

For this study, comparisons between the ANCOVA (measuring posttest differences controlling for pretest differences in the dependent variable) and standardized change score analysis models were compared. Results indicated similar findings; however, the ANCOVA models have greater statistical power to detect an effect (Kenny, 1975) because controlling for pretest differences explains more variance in the outcome than the change score method—a difference in the R^2 between the ANCOVA models and standardized change score of nearly 40% of the total variance explained. While this was the case in this study, it is important to stress that because coefficients were similar for the variables of interest between the change score analysis and the ANCOVA analysis, the results were, in all likelihood, not simply over-estimations of the treatment. Furthermore, the inclusion of covariates, such as relationships and academic motivation, provided additional adjustments to better specify the treatment effect. To that end, only ANCOVA results are reported in this article, although the data and procedures for reproducing the results for both models are available for download.

Six models were specified for each dependent variable. First, a model containing only the demographic and control variables was presented as the base (restricted) model permitting each subsequent model to be compared via F-test. Next, a model containing the indicator for service- learning was added to the restricted model. Finally, four separate models predicted estimates for hours, in-class reflection, out-of-class reflection, and frequency of contact.

Results

This section reports the study results across models for the independent variables of time, reflection, and contact. This study compared these same independent variables to a control group to better understand the magnitude of the effect of the pedagogical variations on civic and academic outcomes. As noted earlier, the dependent variables were standardized to make comparisons across scales. The appendix contains all 36 ANCOVA models (six separate models for predicting six dependent variables). It is also worth noting that the data and statistical procedures were available for replication purposes.

Hours of Service-Learning

The only consistent predictor for the civic engagement outcomes in this study was greater than 30 hours (social values: $\beta = 0.239$, RSE = 0.099; civic attitudes: $\beta = 0.212$, RSE = 0.099; community service self-efficacy: $\beta = 0.243$, RSE = 0.099). Other differences were detectable at 15 or fewer hours, but these results were mixed (social values: $\beta = 0.205$, RSE = 0.098; community service self-efficacy: $\beta = -0.185$, RSE= 0.106). For civic attitudes—the outcome most affected by the pretest non-equivalence—the categories for 16-24 hours ($\beta = .184$, RSE = 0.106) and 25-30 hours ($\beta = 0.264$, RSE = 0.097) were predictive of higher posttest scores.

In-Class Reflection

The in-class reflection models comprised the most consistent predictor of posttest scores across every model except community service self-efficacy. While less frequent in-class reflection was predictive for social values ($\beta = 0.120$, RSE = 0.073) and civic attitudes ($\beta = 0.164$, RSE = 0.072), less frequent reflection had no detectable effect on the posttest scores of any academic measure. Weekly in-class reflection, however, was predictive of positive outcomes for social values ($\beta = 0.336$, RSE = 0.120), civic attitudes ($\beta = 0.343$, RSE = 0.093), general education gains ($\beta = 0.214$, RSE = 0.109), practical competencies gains ($\beta = 0.294$, RSE = 0.103), and personal and social development ($\beta = 0.280$, RSE = 0.107). In each case the model improved, the overall fit improved (e.g., social values: F(2,281) = 4.28, p = .015).

Out-of-Class Reflection

Out-of-class reflection was predictive of positive outcomes for the civic engagement measures. For social values, both weekly ($\beta = 0.232$, RSE = 0.092) and less frequent ($\beta = 0.150$, RSE = 0.078) reflection outside class were detectable. For civic attitudes, a similar pattern was detected for weekly ($\beta = 0.241$, RSE = 0.093) and less frequent ($\beta = 0.184$, RSE = 0.071). Out-of-class reflection was not predictive of any academic measures.

Frequency of Contact with Community Members

After in-class reflection, frequency of contact was the most consistent predictor of outcomes. Infrequent contact had effects that were indistinguishable from zero. Frequent contact, however, was strongly related to positive social values ($\beta = 0.260$, RSE = 0.084) and civic attitudes ($\beta = 0.287$, RSE = 0.090), and weakly related to community service self-efficacy ($\beta = 0.148$, RSE = 0.084) and personal and social development gains ($\beta = 0.131$, RSE = 0.075).

Discussion

It is useful to recall that Mabry's (1998) findings indicated that having frequent contact with beneficiaries, participating in 15 to 19 hours of service-learning activities, engaging in ongoing and summative written reflection, as well as weekly in-class reflection were predictive of outcomes among service-learning students. The present study offers some confirmatory evidence that the findings in Mabry's study are robust in relation to new data and in comparison with students who did not participate in service-learning. The implications of this study increase the overall confidence in the Mabry's original findings. By revealing more supporting evidence for the pedagogical variations that influence student outcomes, service-learning instructors and support staff can be assured that time, contact, and reflection do indeed impact service-learning outcomes.

The findings that were most consistent with Mabry's research focus on the two civic engagement measures borrowed from her original study: social values and civic attitudes. This study confirmed that more than 15 hours of civic engagement produced positive outcomes for students' civic attitudes. However, the only inference that can be drawn from these results is that short-term service-learning experiences may not be effective in producing positive outcomes. In fact, students who engaged in 15 or fewer hours of service-learning activities reported lower community service self-efficacy than students who were assigned no service-learning. This result should be examined further with additional data, though it is consistent with findings from Reeb (2006).

This study confirmed that weekly in-class and out-of-class reflection were effective in producing positive outcomes for social values and civic attitudes. Hatcher, Bringle, and Muthiah (2004) tested Mabry's (1998) assertion that reflection matters to service-learning and focused on the design of service-learning and its relationship to the quality of the learning environment rather than any specific outcomes from service-learning. The current study provided additional evidence that service-learning directly influences outcomes that are desirable for both higher education (Kuh, 2008) and for civic engagement (Bringle, Phillips, & Hudson, 2003).

In this study, frequent contact with service beneficiaries resulted in positive outcomes for service-learners compared to students who did not participate in service-learning. Simons, Russell, Hirschinger-Blank, Williams, and Willis (2009) tested theories of intergroup contact to determine if a culturally based service-learning intervention was more effective in influencing social and personal values related to race and identity. Their findings are consistent with the findings in this study, as frequent contact with community members was shown to influence social values and personal and social development outcomes.

This study also improved upon Mabry's research by introducing better proxy measures for academic outcomes. Mabry noted the following about the findings from the "academic benefit of service-learning"

scale she used:

> Students who served fewer than 15 hours perceived significantly fewer academic benefits than students who served 15 or more hours … [S]tudents who had little or no contact with beneficiaries of their service perceived fewer academic benefits of participating … [S]tudents who participated in weekly in-class reflection perceived greater academic benefits from participating in service-learning than those who did not reflect on their experiences in-class or did so only once or twice in the semester. (p. 40)

This study compared the variations to the control group, so it is difficult to compare the coefficient of one category with another in the same model with a high degree of certainty. However, compared to the control group, only weekly in-class reflection was a significant predictor of positive outcomes across all measures (though weak for general education gains). Frequent contact was weakly predictive of personal and social development gains. The current study did not investigate the kinds of in-class or out-of-class reflection that were included in the original study. Thus, it is possible that specific types of out-of-class reflection were more efficacious than others.

Limitations

This study was limited by its use of quasi-experimental design, which could not control for group non-equivalence, a threat to the internal validity of this study. While academic motivation and relationships were controlled for in the models, the pretest measures of these variables had mixed predictive influence on the posttest outcome. Academic motivation was predictive of the academic measures and relationships for civic engagement measures. These statistical relationships might also have been related to the pedagogical variations themselves, a result of endogeneity (Hart, Donnelly, Youniss, & Atkins, 2007). It is possible that academically motivated students or students who easily form relationships with others are more likely to reflect in class, spend more time in service-learning activities, or interact more frequently with their service beneficiaries, thus resulting in higher posttest scores on the various measures. This effect is illustrated in Figure 1. The only method that can truly overcome this bias is random assignment to treatment conditions.
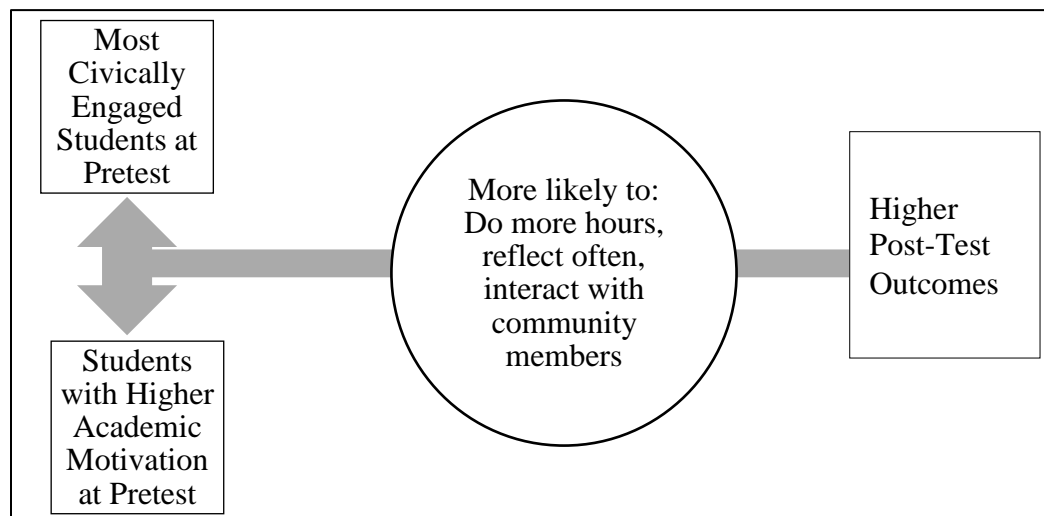

Figure 1. Potential threats of simultaneity in service-learning.

Another limitation present in the current study was low power to detect effects. An alpha level of 0.10

was reported because the use of robust standard errors inflated the error variance for predictors to accommodate the heteroskedastic errors (Gujarati & Porter, 2009). Generally, using robust standard errors enables the analysis of data that otherwise violate the assumptions of linear regression, but it also lowers the statistical power of the tests to correctly avoid both Type I and Type II errors. The challenge of increasing the statistical power of these tests comes in collecting more data.

Future research should also examine the interactive effects of the pedagogical variations. Due to collinearity, it is not possible to conduct such analysis at this time. With additional data, and therefore more variation, it would be possible to model the interactions to answer such questions as "What combination of pedagogical variations is most predictive of higher outcomes?" and to test hypotheses regarding the influences of course design on student outcomes.

Conclusion

This study supports the conclusion that the design of service-learning has implications for how students' attitudes, values, and learning are influenced by the pedagogy. Instructors implementing service-learning must make many decisions about how they will structure the content of their courses, and this study offers additional evidence that service-learning outcomes are related to the length of time spent in community settings, the amount of direct interaction that students have with service beneficiaries, and the frequency of reflection that students have regarding their experiences. Service-learning, like any instructional method, is best implemented when instructors know what methods can be expected to yield outcomes, and this study offers instructors additional confidence that time, contact and reflection matter.

The study also has important implications for the field of service-learning research methodology since it replicates an earlier study showing that, nearly 20 years later, the findings of Mabry's (1998) work are not only consistent with the findings of the present study, but that they transfer to a different, more diverse sample. While both this study and Mabry's study have relatively low power, both detect effects of variations in service-learning pedagogy. These findings reduce the potential for bias and improve the scientific understanding of how service-learning affects students' outcomes.

Finally, this study represents an important advance in service-learning research because it may be the first published study in the field to openly share its data and procedures. As other disciplines increasingly encourage researchers' data to be made publicly available, service-learning researchers should follow suit when it is ethical and reasonable to do so. It is likely that other scholars within the field would willingly and openly share their procedures and data and make them available as appendices to future publications in this journal as well as others. In this way, the principles of replication would be followed more often, and findings would be more transparent.

Author Note

Thomas A. Dahan, Office of Institutional Research, Rutgers University-Camden.

Correspondence

Correspondence regarding this article should be addressed to Thomas A. Dahan, Research Analyst, Office of Institutional Research, Rutgers University-Camden, 406 Penn Street, Camden, NJ 08102-1400. Phone: (856) 225-6969. E-mail: tom.dahan@camden.rutgers.edu

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 Social Science*, *2*, 55-71.

Bringle, R. G. (2003). Enhancing theory-based research on service-learning. In S. Billig & J. Eyler (Eds.), *Deconstructing service-learning: Research exploring context, participation, and impacts* (pp. 3-21). Greenwich, CT: Information Age Publishing.

Bringle, R. G., Hatcher, J. A., & Muthiah, R. N. (2010). The role of service-learning on the retention of first-year students to second year. *Michigan Journal of Community Service Learning, 16*(2), 38-49.

Bringle, R. G., Phillips, M. A., & Hudson, M. (2004). *The measure of service learning: Research scales to assess student experiences*. Washington, DC: American Psychological Association.

Bringle, R. G., & Steinberg, K. (2010). Educating for informed community involvement. *American Journal of Community Psychology, 46(3-4)*, 428-441.

Clayton, P. H., Bringle, R. G., & Hatcher, J. A. (Eds.). (2013). *Research on service learning: Conceptual frameworks and assessment (vol. 2A)*. Sterling, VA: Stylus.

Creswell, J. W. (2008). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Thousand Oaks, CA: Sage.

Delanty, G., & Strydom, P. (2003). *Philosophies of social science: The classic and contemporary readings.* Maidenhead, UK: Open University.

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Thousand Oaks: Sage.

Eyler, J., & Giles, D. E., Jr. (1999). *Where's the learning in service-learning?* San Francisco: Jossey-Bass.

Eyler, J., & Giles, D. E., Jr. (1997). The importance of program quality in service-learning. In A. Waterman (Ed.), *Service-learning: Applications from the research* (pp. 57-76). New York: Routledge.

Gujarati, D., & Porter, D. (2009). *Basic econometrics* (5th ed.). New York: McGraw Hill Irwin.

Hart, D., Donnelly, T. M., Youniss, J., & Atkins, R. (2007). High school community service as a predictor of adult voting and volunteering. *American Educational Research Journal*, *44*(1), 197-219.

Hatcher, J. A., Bringle, R. G., & Muthiah, R. (2004). Designing effective reflection: What matters to service-learning? *Michigan Journal of Community Service Learning*, *11*(1), 38-46.

Indiana University. (2011). *What is the internal consistency of the gains scales? NSSE psychometric portfolio online*. Retrieved from http://nsse.iub.edu/links/psychometric_portfolio

Kenny, D. A. (1975). A quasi-experimental approach to assessing treatment effects in the nonequivalent control group design. *Psychological Bulletin, 82*(3), 345-362.

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter.* Washington, DC: Association of American Colleges and Universities.

Mabry, J. B. (1998). Pedagogical variations in service-learning and student outcomes: How time, contact, and reflection matter. *Michigan Journal of Community Service Learning, 5*, 32-47.

Moely, B. E., & Ilustre, V. (2014). The impact of service-learning course characteristics on university students' learning outcomes. *Michigan Journal of Community Service Learning, 21*(1) 5-16.

Pascarella, E. T. (2007). *Methodological report for Wabash National Study of Liberal Arts Education.* Iowa City, IA: Center for Research on Undergraduate Education.

Reeb, R. N. (2006). The community service self-efficacy scale: Further evidence of reliability and validity. *Journal of Prevention and Intervention in the Community, 32*(1-2), 97-113.

Reeb, R. N., Folger, S. F., Langsner, S., Ryan, C., & Crouse, J. (2010). Self-efficacy in service- learning community action research: Theory, research, and practice. *American Journal of Community Psychology, 46*(3-4), 459-471.

Reeb, R. N., Katsuyama, R. M., Sammon, J. A., & Yoder, D. S. (1998). The community service self-efficacy scale: Evidence of reliability, construct validity, and pragmatic utility. *Michigan Journal of Community Service Learning*, *5*, 48-57.

Simons, L., Russell, B., Hirschinger-Blank, N., Williams, E., & Willis, K. (2009). An exploration of the value of cultural-based service-learning for student and community participants. In B. Moely, S. Billig, & B. Holland (Eds.), *Creating our identities in service-learning and community engagement* (pp.189-214)*.* Greenwich, CT: Information Age Publishing.

Steinberg, K. S., Bringle, R. G., & McGuire, L. E. (2012). Attributes of high-quality research on service

learning. In P. Clayton, R. Bringle & J. Hatcher, *Research on service learning. Volume 2A: Students and faculty* (pp. 27-56). Sterling, VA: Stylus.

Waterman, A. S. (Ed.). (1997). *Service-learning: Applications from the research*. Mahwah, NJ: Lawrence Earlbaum Associates.

Appendix: ANCOVA Models

VARIABLES	Social Values						Civic Attitudes					
	(1) Base	(2) S-L	(3) Hours	(4) In-Class	(5) Out-of-Class	(6) Contact	(7) Base	(8) S-L	(9) Hours	(10) In-Class	(11) Out-of-Class	(12) Contact
Service Learning Variables												
Service-Learning Indicator		0.153**						0.161***				
		(0.064)						(0.060)				
15 or fewer hours			0.205**						0.090			
			(0.098)						(0.088)			
16-24 hours			0.051						0.184*			
			(0.100)						(0.106)			
25-30 hours			0.176						0.264***			
			(0.132)						(0.097)			
Greater than 30 hours			0.239**						0.212**			
			(0.099)						(0.099)			
Weekly in Classroom				0.336***						0.343***		
				(0.120)						(0.093)		
Less Frequent in Classroom				0.120*						0.164**		
				(0.073)						(0.072)		
Weekly outside Classroom					0.232**						0.241**	
					(0.092)						(0.096)	
Less Frequent outside Classroom					0.150*						0.184***	
					(0.078)						(0.071)	
Frequent Contact						0.260***						0.287***
						(0.084)						(0.075)
Infrequent Contact						0.054						0.090
						(0.088)						(0.083)
Control Variables												
Pre-test of Dependent Variable	0.602***	0.593***	0.623***	0.603***	0.609***	0.609***	0.645***	0.633***	0.631***	0.609***	0.617***	0.614***
	(0.050)	(0.050)	(0.055)	(0.054)	(0.055)	(0.054)	(0.053)	(0.053)	(0.059)	(0.063)	(0.061)	(0.061)
Relationships (Pre Test)	0.044	0.041	0.049	0.082*	0.086*	0.086*	0.066	0.064	0.065	0.103**	0.101**	0.093*
	(0.043)	(0.042)	(0.046)	(0.046)	(0.046)	(0.048)	(0.046)	(0.045)	(0.050)	(0.050)	(0.050)	(0.051)
Academic Motivation (Pre Test)	0.082	0.071	0.040	0.041	0.044	0.026	0.064	0.056	0.020	0.022	0.036	0.028
	(0.060)	(0.059)	(0.068)	(0.062)	(0.064)	(0.065)	(0.064)	(0.064)	(0.072)	(0.069)	(0.068)	(0.068)
Demographic Controls?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.188	0.114	0.048	0.050	0.123	0.096	-0.033	-0.103	-0.028	-0.028	-0.004	0.000
	(0.153)	(0.158)	(0.174)	(0.178)	(0.179)	(0.181)	(0.129)	(0.130)	(0.134)	(0.132)	(0.135)	(0.133)
Observations	360	360	296	296	297	293	402	402	331	331	332	328
R-squared	0.425	0.435	0.456	0.466	0.458	0.464	0.432	0.442	0.447	0.457	0.451	0.459

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Appendix: ANCOVA Models

VARIABLES	Community Service Self-Efficacy						Gains in General Education					
	(13) Base	(14) S-L	(15) Hours	(16) In-Class	(17) Out-of-Class	(18) Contact	(19) Base	(20) S-L	(21) Hours	(22) Reflection	(23) Reflection	(24) Contact
Service Learning Variables												
Service-Learning Indicator		0.084						-0.049				
		(0.066)						(0.069)				
15 or fewer hours			-0.185*						-0.067			
			(0.106)						(0.118)			
16-24 hours			0.152						-0.033			
			(0.106)						(0.135)			
25-30 hours			0.171						0.068			
			(0.125)						(0.110)			
Greater than 30 hours			0.243**						-0.030			
			(0.099)						(0.121)			
Weekly in Classroom				0.140						0.214*		
				(0.130)						(0.109)		
Less Frequent in Classroom				0.081						-0.088		
				(0.073)						(0.084)		
Weekly outside Classroom					0.133						-0.010	
					(0.116)						(0.137)	
Less Frequent outside Classroom					0.096						-0.027	
					(0.075)						(0.079)	
Frequent Contact						0.148*						0.031
						(0.084)						(0.085)
Infrequent Contact						0.030						-0.091
						(0.096)						(0.107)
Control Variables												
Pre-test of Dependent Variable	0.682***	0.679***	0.679***	0.675***	0.674***	0.672***	0.399***	0.398***	0.389***	0.377***	0.381***	0.386***
	(0.050)	(0.050)	(0.051)	(0.054)	(0.053)	(0.053)	(0.056)	(0.056)	(0.060)	(0.056)	(0.059)	(0.061)
Relationships (Pre Test)	0.103**	0.101**	0.071	0.130**	0.126**	0.129**	0.039	0.041	0.022	0.050	0.048	0.050
	(0.049)	(0.049)	(0.053)	(0.054)	(0.053)	(0.055)	(0.051)	(0.052)	(0.057)	(0.058)	(0.058)	(0.059)
Academic Motivation (Pre Test)	0.013	0.008	-0.017	-0.038	-0.029	-0.042	0.286***	0.289***	0.314***	0.284***	0.300***	0.295***
	(0.067)	(0.067)	(0.074)	(0.072)	(0.071)	(0.073)	(0.065)	(0.065)	(0.073)	(0.069)	(0.071)	(0.071)
Demographic Controls?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	-0.161	-0.198	-0.064	-0.090	-0.049	-0.081	0.200	0.218	0.408**	0.363*	0.359*	0.394**
	(0.140)	(0.144)	(0.160)	(0.157)	(0.152)	(0.163)	(0.158)	(0.161)	(0.194)	(0.186)	(0.191)	(0.189)
Observations	402	402	331	331	332	328	388	388	322	321	322	318
R-squared	0.505	0.507	0.520	0.510	0.513	0.509	0.321	0.322	0.342	0.351	0.339	0.343

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Appendix: ANCOVA Models

VARIABLES	Gains in Practical Competencies						Gains in Personal and Social Development					
	(25) Base	(26) S-L	(27) Hours	(28) In-Class	(29) Out-of-Class	(30) Contact	(31) Base	(32) S-L	(33) Hours	(34) In-Class	(35) Out-of-Class	(36) Contact
Service Learning Variables												
Service-Learning Indicator		0.060						0.083				
		(0.063)						(0.061)				
15 or fewer hours			0.040						-0.004			
			(0.105)						(0.099)			
16-24 hours			0.058						-0.001			
			(0.124)						(0.112)			
25-30 hours			0.191*						0.159			
			(0.106)						(0.113)			
Greater than 30 hours			0.016						0.074			
			(0.105)						(0.107)			
Weekly in Classroom				0.294***						0.280***		
				(0.103)						(0.107)		
Less Frequent in Classroom				-0.022						-0.009		
				(0.075)						(0.070)		
Weekly outside Classroom					0.103						0.076	
					(0.122)						(0.106)	
Less Frequent outside Classroom					0.045						0.056	
					(0.073)						(0.072)	
Frequent Contact						0.110						0.131*
						(0.078)						(0.075)
Infrequent Contact						-0.008						-0.034
						(0.097)						(0.091)
Control Variables												
Pre-test of Dependent Variable	0.380***	0.379***	0.360***	0.356***	0.357***	0.349***	0.465***	0.465***	0.439***	0.429***	0.430***	0.408***
	(0.056)	(0.056)	(0.060)	(0.059)	(0.062)	(0.062)	(0.060)	(0.060)	(0.069)	(0.065)	(0.066)	(0.065)
Relationships (Pre Test)	0.045	0.044	0.041	0.040	0.034	0.046	0.045	0.043	0.047	0.072	0.068	0.081*
	(0.047)	(0.047)	(0.052)	(0.052)	(0.052)	(0.053)	(0.043)	(0.043)	(0.048)	(0.047)	(0.048)	(0.048)
Academic Motivation (Pre Test)	0.177***	0.173**	0.157**	0.166**	0.187**	0.186**	0.139**	0.133**	0.118**	0.087	0.106*	0.112*
	(0.067)	(0.067)	(0.071)	(0.071)	(0.073)	(0.073)	(0.055)	(0.055)	(0.059)	(0.057)	(0.058)	(0.058)
Demographic Controls?	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Constant	0.126	0.103	0.226	0.208	0.226	0.221	-0.109	-0.141	0.023	-0.046	-0.027	-0.021
	(0.146)	(0.148)	(0.170)	(0.166)	(0.172)	(0.169)	(0.134)	(0.136)	(0.155)	(0.142)	(0.148)	(0.146)
Observations	388	388	322	321	322	318	390	390	323	322	323	319
R-squared	0.283	0.285	0.282	0.303	0.283	0.286	0.327	0.330	0.308	0.330	0.311	0.315

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$