© Journal of Higher Education Outreach and Engagement, Volume 27, Number 2, p. 161, (2023)
 eISSN 2164-8212

Developing the SLQAT (Service-Learning Quality Assessment Tool), a Quantitative Instrument to Evaluate Elements Impacting Student Outcomes in Academic Service-Learning Courses

Paul H. Matthews, Isabel Lopez, Laurel E. Hirt, Shannon O. Brooks, and Andrew Furco

Abstract

Given the diversity of settings and courses representing academic service-learning practice, a standardized, quantitative instrument to rate the quality level of course design and implementation is needed to optimize educational outcomes for participating students. This article describes a 5-year, multi-institutional process developing the *Service-Learning Quality Assessment Tool (SLQAT)*, a quantitative diagnostic composed of 28 "essential elements" known to promote positive student outcomes in postsecondary service-learning. We discuss the selection and operational definitions for these elements, the assumptions and decisions behind the development of the instrument, the use of expert feedback to develop baseline weights representing the relative importance of each element's contribution, the creation of rating levels representing element quality, and the development of protocols for the instrument's scoring and uses. We also reflect upon the challenges of attempting to create a broadly applicable instrument and share plans for additional piloting as well as recommendations for research and practice.

Keywords: service-learning quality, quantitative instrument development, student learning outcomes, service-learning assessment, course quality rubric

Academic service-learning—a pedagogy in which students' course knowledge is applied and shaped through collaboration with and service to community partners—is intentionally not a one-size-fits-all proposition; "indeed, no two service-learning activities are alike" (Furco, 2003, p. 13). Service-learning experiences are molded by the particular academic and community contexts in which they operate, and, in turn, are designed for particular outcomes and purposes across different stakeholders (Langhout et al., 2023; this issue). For example, even when focusing only on research investigating students, service-learning has consistently been found to achieve a broad range of positive outcomes, such as improved academic achievement (e.g., Kuh, 2008; Warren, 2012), enhanced personal and social development (e.g., Brandenburger, 2013), increased civic responsibility (e.g., Conway et al., 2009; Yorio & Ye, 2012), retention and persistence toward graduation (e.g., Bringle et al., 2010; Lockeman & Pelco, 2013; Mungo, 2017; Provencher & Kassel, 2017; Song et al., 2017), and even postgraduation employment benefits (e.g., Matthews et al., 2015), to name only a few.

We also know that for achieving these positive student outcomes, course quality matters (Billig, 2009; Billig et al., 2005; Eyler & Giles, 1999; Kuh, 2008; Mabry, 1998). Indeed, research studies have identified a number of key practices as fundamental to the integrity and quality of service-learning courses, both in K-12 and higher education settings. Although an exhaustive review of the literature on service-learning best practices is beyond the scope of this article,

dozens of different elements have been identified, summarized, or hypothesized by past scholarship as having impact on student outcomes (e.g., Botelho et al., 2020; Eyler & Giles, 1999; Heffernan, 2001; Jacoby, 2015; Steinke et al., 2002; Waterman, 2003). Additionally, service-learning courses that implement more of these practices identified as essential elements are more likely to result in positive outcomes for students (Celio et al., 2011).

However, "[i]t is simplistic to believe that following general principles of good practices in service-learning will affect all outcomes equally" (Steinke et al., 2002, p. 77). In addition, these practices are not incorporated across service-learning courses to the same degree (if at all). From course to course, service-learning practice can vary across a range of variables, representing differences in course design, partnerships, student experience, and instructor and institutional characteristics (Bringle et al., 2013; Furco, 2003; Heffernan, 2001; Roldan et al., 2004; Waterman, 2003). Even a cursory consideration of logistical possibilities—for instance, the amount of service provided, the service type (direct, indirect, nondirect), the degree to which service activities are integrated with the academic curriculum, students' preparation for service activities, and frequency and type of reflection—suggests many ways that courses vary. Experienced instructors also recognize that even for the "same" course, the specific implementation of the pedagogy is mutable from one semester to the next and among individual students' experiences. As an example, the engagement of students in reflection and analysis about the academic learning and societal impact of their work is considered an essential, undisputed best practice of service-learning (e.g., Eyler & Giles, 1999; Hatcher et al., 2004; Jacoby, 2015). Yet, even in our own intrainstitutional and cross-institutional analyses, we have found tremendous variation in what such reflection looks like. At the University of Georgia, for instance, among courses designated as service-learning, students report taking part in reflection between 0 and 20 times per semester (mean reported for fall semester 2019 was 8.4 instances), through as many as 10 different types (mean, 3.5) of reflective activities in their course. Similar variations in practice are found among service-learning courses at the University of Minnesota.

Critically, our field lacks quantitative instruments with which to capture differences or track the presence of key practices, much less the nuances of implementation quality (e.g., Bailis & Melchior, 2003; Botelho et al., 2020; Shumer, 2003). The diversity of service-learning practice poses challenges and limitations to conducting studies of service-learning with fidelity (Furco, 2003), especially for larger scale, institutional, and multisite research (Bailis & Melchoir, 2003). Most such studies, including ones conducted by members of our research team, end up simply having to categorize courses in a binary, as "service-learning" or "not service-learning" (e.g., Matthews et al., 2015; Song et al., 2017; Wilder et al., 2013), which runs the risk of oversimplification, obscuring important details and practices within the "service-learning" category.

Consistent, quantitative measurement of the presence and quality of best practices would better allow for statistical comparisons and more nuanced analyses across service-learning experiences, courses, and programs. Although some consensus exists on what these quality components of service-learning are, there is no standardized, quantitative instrument available that allows practitioners or scholars to assess the extent to which a course incorporates these key elements of high quality practice. Existing instruments are primarily qualitative, and/or are focused on only a few key components or particular disciplines. For instance, Shumer (2003) reported on a 3-year project to develop a self-assessment instrument for service-learning practitioners in K-12 settings (*The Quintessential Elements of Service-Learning*), with 23 statements in five domains; however, this instrument was designed primarily for program improvement, allowing for self rating of each only as "weak," "needs work," or "strong." Jenkins and Sheehey (2011) developed a staged "checklist for planning, implementing, and evaluating service-learning" (p. 54); their instrument is intended for course design, and does not include ratings. Similarly, Welch's (2010) O.P.E.R.A. model provides a planning framework with five key practices, but is not suitable for research. IUPUI's "taxonomy for service learning courses" (Hahn et al., 2016) details six important aspects of service-learning course design, each with three levels of implementation, but does not purport to address all quality elements, nor does it provide any sense of relative im-

portance of these components. Kieran and Haack (2018) developed a rubric "to evaluate course syllabi for quality and evidence-based indicators of [service-learning] components as found in the literature" (p. 42). Their PRELOAD rubric includes dimensions of partnership, reflection, engagement, logistics, objectives, assessment, and definition of service-learning as of importance, with scoring possibilities of "excellent," "satisfactory," and "developing"; however, this rubric is still oriented toward syllabus design, rather than actual implementation. Stokamer (2018) led a group at her university to develop a set of 10 Principles of Quality Academic Civic Engagement (PQACE) based in "the S-LCE literature, best practices, and personal experience" (p. 224) and geared toward their specific university context. Botelho et al. (2020) used student and faculty surveys and syllabi to determine a set of eight components of service-learning quality in STEM courses across the California State University system. These included both composite measures ("reflections," "values focus," "collaboration with community," "addressing community need," "linked to academic content," and "communication with community") and single-item components ("service-learning preparation" and "linked to learning objectives"), each of which could be rated on a scale of 1 to 4 (or 5) based on review of STEM syllabi and postparticipation student surveys.

In this article, we describe a 5-year, multi-institutional initiative intended to address the challenge and need for a standardized, quantitative, and scorable rating instrument focused on service-learning implementation and design. Below, we describe the iterative and cyclically reflective process (e.g., Kolb, 1984) of conceptualizing, developing, piloting, redesigning, weighting, and offering an instrument to the service-learning community, in order to allow researchers to evaluate more consistently the impacts of different essential elements of service-learning on student outcomes. We also reflect upon some of the challenges and decision points in the process, potential uses (and misuses) of such an instrument, and next steps for both our research team and the field.

Developing a Standardized Rating Instrument to Measure Service-Learning Quality

Purpose and Assumptions

The instrument—the *Service-Learning Quality Assessment Tool (SLQAT)*—was designed to address the need for a quantitative, comprehensive tool that allows for consistent and differentiated ratings of multiple key aspects associated with high quality design and implementation of service-learning courses in higher education, specifically oriented to student academic learning outcomes. The original impetus for its design lay in the larger, federally funded research program examining the impact of various community engagement practices on underrepresented undergraduates' educational success. In investigating service-learning course impact on student learning and educational success, members of the research team were interested in controlling and accounting for the quality of students' service-learning experience. Specifically, they sought to find a means to establish for each service-learning course a quantitative score that indicated the level of quality, based on the course's inclusion of service-learning best practices.

Although the SLQAT was born out of a study focused on outcomes for underrepresented students, the researchers conceptualized and developed the SLQAT as a more generally applicable research tool appropriate for all types of service-learning courses and all student populations. In addition, as is discussed further below, this measurement tool has broad utility beyond conducting research. For example, it can be used as a guide to conduct institutional reviews or approvals of service-learning courses. Faculty members can also use the tool when developing their own courses to ensure the inclusion of the essential elements of service-learning. Administrators can use the instrument as part of institution-wide self studies designed to identify the strengths and weaknesses of their institutions' service-learning and community engagement efforts.

Several assumptions guided the process and development of the instrument, resulting in choices of both what elements to include or exclude and how to orient, structure, and use the SLQAT. These assumptions and choices related primarily to three areas: definition of the service-learning context/

setting, selection of essential elements, and identification of data sources for scoring.

Service-Learning Context for Application

Regarding the context of the instrument, the SLQAT is based on best practices that pertain to service-learning in postsecondary (i.e., college/university) course settings. Following Bringle and Hatcher's (1995) characterization of service-learning as "course-based, credit bearing," the instrument is also designed strictly for evaluating *curricular* service-learning, not cocurricular experiences. In addition, service-learning is assumed to be a required (rather than optional) component of the course. Although other stakeholder outcomes (e.g., impact on the community) are key considerations for service-learning, this tool is focused tightly on *student learning outcomes* and the practice elements that influence them. Finally, the instrument aspires to be *universal*—relevant to and usable in all types of service-learning courses, regardless of discipline, length of engagement, service activity type (direct, nondirect, or indirect service), institutional type, location, or other contextual variables (Furco, 2003).

Selection of Essential Elements

Several key principles guided choices by the research team on what to include as "essential elements" (Billig et al., 2005; Botelho et al., 2020) in the SLQAT. First, in line with the above, individual elements should be broadly (or even universally) *applicable* across the range of disciplines, settings, and levels represented in service-learning coursework. Second, each element is assumed to be *essential*, in that research and/or practice suggest that it contributes tangibly and independently to the overall quality of service-learning student outcomes. Thus, any course that does not include all these elements is hypothesized to be less effective at bringing about positive student outcomes, in the same way that excluding key ingredients in a recipe will not result in as satisfactory a culinary outcome.

However, not all elements are assumed to contribute equally to service-learning quality (Steinke et al., 2002); for instance, in the previous analogy, the impact of leaving meat out of a pot roast recipe is likely more impactful than omitting celery. In the SLQAT, this is represented through differing base score values or weights that represent each element's level of hypothesized importance, as described later. In addition, elements should be able to be *substantiated*; each element should be clearly defined so that its absence, presence, and level of implementation can be consistently and definitively ascertained during rating. Finally, we acknowledge that a host of other factors likely also influence the quality of service-learning courses and implementation (e.g., faculty teaching experience, size of the course, length of term, students' prior experience with service-learning, access to transportation, community and institutional characteristics, etc.). However, as such factors typically cannot be adjusted at the course level or are out of the instructor's control, selection of elements for the SLQAT was oriented toward those that are *responsive* to the instructor's influence.

Scoring Assumptions

Other assumptions relate to the use and scoring of the SLQAT (further described later). For instance, scoring is based on a particular *instantiation* of a course (i.e., a product of a given semester and instructor, rather than a generic "master syllabus"), and the course is assumed to have been taught prior to scoring. Additionally, information contained in the data sources analyzed (such as the syllabus) is assumed to *represent actual practice* in the delivery of the course, and thus to be valid for determining the presence or absence of each element. Finally, in terms of *construct validity*, higher scores on the SLQAT are assumed to represent a higher quality of service-learning course implementation, which in turn is assumed to produce more positive outcomes for students.

Initial Conceptualization of the SLQAT

Instrument development was an iterative process from 2016 to 2021, engaging multiple stakeholders. The primary research team consisted of administrative faculty, staff, and graduate students at both the University of Georgia and the University of Minnesota. Key members of the team have decades of experience in service-learning administration, research, and teaching. The team met approximately monthly, typically virtually, over a 5-year period, with frequent emails and shared online documents and drafts, as well as periodic in-person work sessions. Team members also shared drafts and consulted with other researchers and practitioners in the service-learning/community engagement field at conferences and directly, throughout the process.

The initial instrument development began by brainstorming an intentionally large list of potential best practices for service-learning, based on the research team members' understanding of research and practice, resulting in nearly 50 potential elements for consideration. These potential elements were discussed and consolidated, following the principles and assumptions guiding the project as outlined above. Each potential essential element was given a short title and a short description, then elements were grouped (and regrouped) thematically into a subset of categories or "dimensions" and numbered for ease of reference. See Appendix for a full list of element titles and short descriptions. A full version of the tool (Furco et al., 2023) is published in this special issue. Early versions considered as many as 38 prospective essential elements, representing different dimensions (learning, service, student, faculty, community, structural, program improvement, institutional policies, etc.).

Weighting Essential Elements

Next, an initial weighting by a subset of the research team was performed for 36 initial elements, with ratings assigned as 1 (*slightly important*), 2 (*somewhat important*), or 3 (*very important*) to student learning outcomes. These individual ratings were compared and discussed, with sustained, deep discussion on wording, relevance, and importance. Means and standard deviations across the individual ratings were reviewed, and any element scored with more than a 0.5 standard deviation in mean (i.e., not rated the same by two or more of the five raters) was discussed or modified to achieve consensus. The revised mean rating served as an initial quantitative representation of the relative importance of that element, but more importantly, the process provided a continuous review of the clarity (conceptual as well as descriptive) of the instrument's elements and of the assumptions guiding its development.

During the next year, the essential elements were winnowed down as the process of piloting with real courses began. The intent of this pilot process was to ensure elements were clearly defined and operationalized, applicable to different types of service-learning, and sufficiently distinct from each other. Thus, some elements that were initially posited to impact student learning were removed when they were deemed difficult to substantiate based on the review of submitted course materials. Other elements were removed or reworked based on the realization that there would likely not be any course-to-course variability within the same institution (e.g., "institutional climate for service-learning") or as insufficiently focused on service-learning (e.g., "syllabus goals, expectations, requirements and assessment criteria clearly stated").

A second round of element weighting was performed in late 2016 with a revised set of 30 elements and weights. Seven raters from the research team scored each element, with subsequent in-depth group discussion on each element. Any elements with a standard deviation exceeding 0.5 were extensively discussed, and outlier ratings were voluntarily modified to fall within this parameter. Next, the mean scores of the finalized seven ratings were tallied to create an initial "base score" (ranging from 1.29 to 3.0). At the 2016 meeting of the International Association for Research on Service-Learning and Community Engagement (IARSLCE), the instrument was presented and session participants were invited to submit their own individual ratings for each element via a Qualtrics survey on the same scale (0.5 to 3.0). Comparing the IARSLCE attendees' means for each element with the research team's initial means showed that 23 of these 30 elements were rated with less than 0.5 difference (i.e., one scalar point) in either direction, suggesting that element score ratings could be "crowdsourced" with results similar to the more extensively deliberated ratings assigned by the research team. IARSLCE raters also were invited to share feedback on the instrument and the elements, which were reviewed and discussed by the research team, leading to additional modifications.

Additional Piloting and Feedback

The revised set of 30 elements was next piloted more broadly by the research team in spring 2017 with a purposive convenience sample of four courses (two from the University of Georgia and two from the University of Minnesota). For this round of the instrument's development, a series of quality level statements was created in order to operationalize or describe "baseline" level implementation, as well as "below baseline" and "above baseline" levels; these latter categories furthermore had two possible levels of quality within each descriptor, allowing five possible rating levels. The research team's mean scores for each element

were used as the baseline value of each element, then converted into five categories of weights: 20% below baseline, 10% below baseline, baseline, 10% above baseline, or 20% above baseline (see Figure 1). The service-learning courses for this scoring were all established courses at the two universities, each at the 3000 level, and were intended to provide diversity in discipline, quality, and service type (two direct service, two indirect service), to assess how well the instrument could be used in differing course settings. They included a small-group communication studies course in which students collaborated with nonprofits on a range of projects, then reflected on how they applied group work strategies, communication, and leadership; an online adolescent development course in which students provided peer mentoring for adolescents around the world through an online collaboration; a technical/professional writing course in which students developed written project deliverables for a set of community partner organizations; and an education course engaging preservice teachers in working with youth in educational settings and blogging about their experiences.

As part of this pilot and the challenges that emerged while scoring these courses, our team recognized that additional information beyond just the syllabus would likely be needed to definitively score the presence or absence of all elements. Discussion and reflection around points of disagreement or divergent interpretation of elements led to additional edits in the language, organization, and wording of elements over the next several months, and two more elements were removed or consolidated (e.g., "connection to broader socially relevant issues" was merged with "societal issues learning"). In late 2017, another round of pilot scoring using 28 elements was conducted (with the same technical writing course), resulting in further refinement of the language describing and naming the elements.

In order to engage and obtain feedback from the broader scholarly community, additional workshops and presentations of the instrument were made at numerous national and international venues from 2016 to 2019, including IARSLCE, the Engagement Scholarship Consortium, the Gulf-South Summit on Service-Learning and Civic Engagement Through Higher Education, Campus Compact conferences, and international research gatherings. At each venue, we solicited participant feedback related to the instrument and rating process, and promoted the opportunity to participate in future pilots.

Methodology for Restructuring Baseline Weights of Elements

In 2019, the research team reevaluated the prior baseline weighting of elements. We wished to address concerns that subsequent editing of the instrument had potentially shifted the element descriptions since the initial weighting, as well as addressing concerns and feedback about the meaningfulness of differentiating weights to the second

Figure 1. Sample Essential Element With Quality Statements, Implementation Levels, and Weighting

	Element #1: Articulation of Service-Learning in Syllabus			
Short Description	*Service-learning is articulated and integrated in the course design and syllabus*			
Quality Statements	**Is there evidence in the syllabus of a service-learning experience within the course design and/or the course expectations?**	Element is absent based on existing evidence.	While the SYLLABUS or ancillary documents mention a service-learning experience, this is underdeveloped, unclear, not relevant, or not integrated into the rest of the course.	The SYLLABUS articulates and describes a relevant service-learning experience as part of the course.	The SYLLABUS clearly explains the scope, relevance, and purpose of the service-learning experience, and how it is integrated into the course, with appropriate details.
Implementation Level:	**Absent**	**Below Baseline**	**Baseline**	**Above Baseline**
Weighted Element Score:	0	5.6	7.5	9.4
Evidence/Notes:				

Note. See Appendix for full list of elements and short descriptions.

decimal place, which suggested a level of precision beyond our actual methodology. The range of possible scores, the appropriate level of precision, and the overall size of the weights were extensively discussed. Discussion included issues such as the merits of a 3-point, 5-point, or other scale; the likelihood that a score such as 2.13 was or was not meaningfully different from a score such as 2.33; and the impact of higher versus lower possible weights on overall scoring when some elements are scored absent, to name a few.

Ultimately, our research team decided to solicit additional expert feedback from the larger scholarly community. In 2020 we emailed invitations nationally and internationally on relevant email lists and through direct invitations to service-learning scholars and practitioners to independently quantify the posited value of each element, with no preconceived basis or provision of our own research team's prior scores. This process invited raters to read each of the 28 elements and its short description, then to assign a weight ranging from 1 to 9 to allow for greater nuance or spread, based on the influence of the given element on student learning outcomes. Participants were also asked to provide feedback on the validity, comprehensiveness, and wording of the instrument, and to self-rate their level of expertise and experience in service-learning teaching and research.

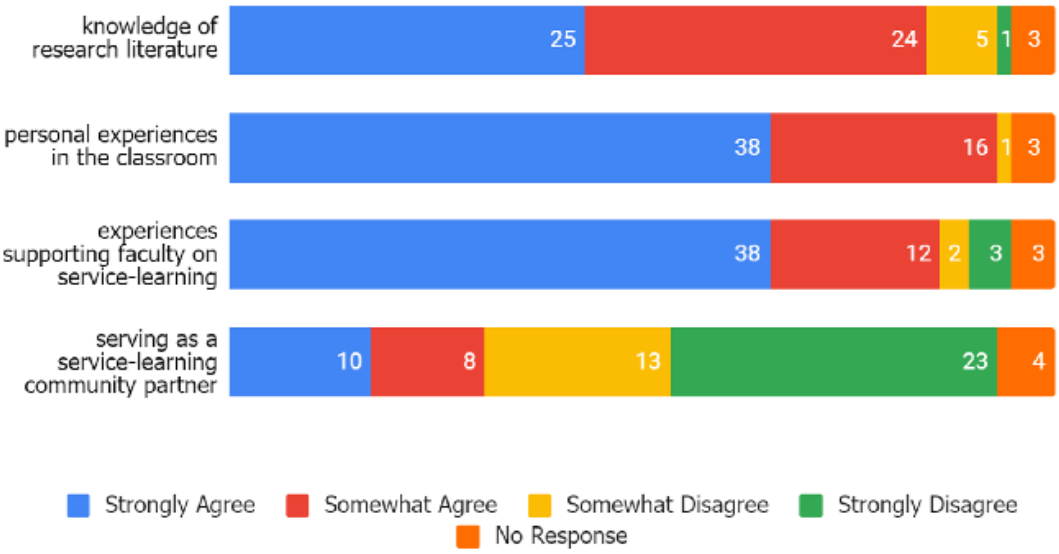
Some 65 responses were recorded through both Qualtrics and GivePulse platforms. Responses from members of our research team and from respondents who did not complete the weighting matrix, as well as a sole respondent who described their "knowledge of service-learning research and practice" as "novice" level, were eliminated from the data set. This step resulted in a final pool of 58 respondents, who represented instructional faculty, administrators, and other roles, primarily in higher education settings (see Table 1). Respondents were mostly from public (n = 24) and private (n = 17) institutions in the United States (representing 29 states), about half of which held the Carnegie community engagement elective classification, as well as from eight private and public universities in seven other countries. These respondents also explained the basis for their ratings, as shown in Figure 2.

The survey also provided respondents the opportunity to propose "any service-learning course design elements that are missing which impact student learning." All comments (n = 27) were carefully reviewed, categorized, and assessed in light of the same assumptions and guidelines used for the extant elements. Most suggestions were already represented in extant elements, though not always clearly articulated in the short description of the elements provided to raters (e.g., Element #14:

Table 1. Self-Reported Characteristics of Rating Respondents

Role		
Instructional faculty	21	
Administrator	32	
Other role	5	
Institutional Affiliation		
Higher education	52	
Non higher education	3	
No institutional affiliation	3	
Experience		
	Yes	*No*
Has taught service-learning courses	53	5
Has published service-learning research	36	22
	Advanced	*Intermediate*
Level of service-learning knowledge	36	22

Figure 2. Basis for SLQAT Elements Rating Responses



Appropriateness of Service Activities for Students—The service activities are contextually appropriate for students' level of skill/knowledge/experience). In some cases we clarified or strengthened them further in the SLQAT's quality level statements (see Figure 1). Other suggestions were not applicable to the full gamut of service-learning experiences (e.g., were relevant only for a certain discipline, or only for direct-service activities, etc.). One tangible change recommended by an expert rater resulted in renaming one element (from "reciprocity" to "mutual benefit") to be more in line with how the element is described in the instrument and supporting literature.

Although invited to rate these elements on a scale of 1 to 9, respondents' ratings of the baseline weights showed that they generally considered all the elements to be highly impactful on student learning outcomes, with an overall mean of 7.42 (SD, 1.48) and individual mean element weights ranging from 5.83 to 8.55 on the 9-point scale. This reinforced the assumption that these elements are indeed essential to service-learning. A further comparison of the ratings assigned by respondents who self-identified as having an "advanced" versus "intermediate" level of service-learning experience showed that the more expert raters identified the elements as even more impactful on average (a summed mean difference of 12.08 across the set of 28 elements). Because these differences in mean group ratings were statistically significant ($t(54) = 2.72$, $p < .01$), we decided to use the ratings by the "advanced" group only ($n = 36$), in order to maximize the expertise of the rater pool. Furthermore, because minor differences of tenths or hundredths of points seemed unlikely to represent meaningful variation of importance across elements, mean scores for each element were rounded to the nearest 0.5, resulting in final weights ranging from 6.0 to 9.0 with an approximately bell-curved distribution (Table 2). The spread of these base weights suggests that the lowest rated element could be considered about two thirds as impactful on student learning outcomes as the highest rated one. Additionally, with these 28 baseline weights summing to 212.5, any element marked as "absent" would reduce the summed total by about 7.6 points on average.

Assigning Implementation Quality Levels

In line with the goal of creating an instrument responsive to difference, each element was intended to be scorable on a range of levels of implementation quality, with concomitant differences in the weight assigned based on the hypothesized importance of the element's contribution to student learning outcomes. Earlier iterations of the instrument had proposed five categories of implementation quality, with varying values assigned to each level. However, pilot rater feedback showed that distinctions within the upper two (i.e., +10% vs. +20%) as well as the lower two (i.e., -10% vs. -20%) gradations were not able to be made con-

Table 2. Distribution of Baseline SLQAT Element Weights

Baseline Weight	Number of Elements (n = 28)	Elements With This Weight
6.0	1	#9
6.5	1	#8
7.0	5	#4, #7, #18, #19, #20,
7.5	10	#1, #5, #10, #13, #14, #17, #21, #23, #25, #26
8.0	8	#3, #6, #11, #12, #22, #24, #27, #28
8.5	2	#15, #16
9.0	1	#2

sistently. Therefore, despite the potential loss of nuance, we opted to enhance usability and consistency and consolidated possible ratings of quality to three levels (i.e., "below baseline," "baseline," and "above baseline"; see Figure 1).

Our next consideration was determining the appropriate spread to quantify these levels of quality within each element. We considered whether "above baseline" or "below baseline" should best be operationalized as reducing or enhancing the value of the baseline weight by 10%, or by 25%, 50%, or some other amount. We also discussed at length the benefits and challenges for different scale points and categories; for instance, whether to make the ratings represent a continuous variable (i.e., with a true zero for absent elements and consistent intervals between zero and each of the subsequent three quality levels), which could have advantages in terms of possible statistical procedures applied to the scores. In reviewing the element quality categories, we concluded that we were not operationalizing each of the three quality categories as representing consistent quantity or level of difference between quality categories, suggesting that these rating categories are more likely to represent ordinal-level points.

We also considered the practical interpretive implications for overall summed scores as described below using these possible spreads of ratings. Higher percentage values would raise the stakes for accuracy of rating across the three implementation levels, since moving from one quality category to the next in a broadly spread scoring scheme would have greater impact on the overall summed score than in a scenario with relatively less change in scores based on quality level. Analogously, on a ±10% plan, a course would have to score above baseline on about 10 additional elements for every element missing in order to receive a summed total quality score equivalent to that of a course with all elements rated as present at the baseline level of quality. Conversely, on a ±50% scoring plan, a course with a single missing element and two elements rated above baseline would receive a summed total score about equal to a course with all "baseline and present" scores.

In the absence of compelling data to substantively inform these decisions, our team agreed that the element ratings are likely ordinal-level variables, and opted for an intermediate level of impact by assigning ±25% as the variation from baseline for the quality categories. Ratings for particular elements present in a course therefore might range from 4.5 ("below baseline" for Element #9) to 11.3 ("above baseline" for Element #2). This broad set of possible ratings thus reflects hypothesized differences in both importance (baseline weight) and implementation quality of these essential elements.

Using the Service-Learning Quality Assessment Tool

The most current iteration of the SLQAT (Furco et al., 2023) consists of 28 essential elements, numbered and grouped for convenience into five conceptual dimensions (course design, learning, student, instructor, and community partner/partnership). Each essential element in the SLQAT has a title, a short description, a question to guide determination of evidence of the element's presence or absence in the data sources, and three levels of descriptive text with corresponding implementation quality categories. As described previously, the SLQAT includes a corresponding, underlying baseline weight (numerical value) for each element, repre-

senting the hypothesized importance of that element's contribution to service-learning course quality and implementation. In each element, the three quality categories help raters determine how well the element is put into practice: whether best described as presenting at baseline level (present with adequate implementation, scored at the base weight for the element), below baseline (partial or inadequate implementation, scored at 25% below the base weight), or above baseline (exemplary in implementation quality, scored at 25% above the base weight for that element). Each element rating block also includes an "evidence/notes" section where a rater may list comments, questions, or notes on what evidence their rating draws upon. Scoring is based on the overall evidence provided about the course, as described in the following section.

Course Evidence and Scoring Guidelines

The scoring process for a given course is intended to be based on a review of both *foundational* and *supplemental* data sources. The foundational sources for scoring the SLQAT are those deemed essential for rating, and include the course syllabus and course-specific materials provided to students (e.g., assignment guidelines not incorporated into the syllabus; student contracts for service-learning; information about community partners, placements, or projects; pertinent service-learning handouts from the institution's service-learning office). Based on pilot rating to date, foundational materials alone typically do not provide sufficient evidence to determine the presence/inclusion of all of the SLQAT's elements. Thus, using one or more supplemental data sources in the rating process is likely necessary to help enhance the accuracy and confidence of ratings. Supplemental data sources may include items such as interviews with or statements from the instructor; information from the campus service-learning office, the community partner, and/or students who took the course; sample deliverables from the service-learning activity; student reflections; and similar sources.

Additionally, our pilot testing suggests that at least two raters should use this instrument to independently rate a given course. Multiple raters can enhance objectivity and reduce potential rater error, thus strengthening the reliability of the scoring process, especially when discussion of program elements is included (cf. Shumer, 2003, p. 154). We recommend that each rater carefully review the initial course materials and independently score each element in the SLQAT, noting evidence supporting each rating. For elements where the data provided do not allow the rater to decide if the element is truly absent, a preliminary indication of "insufficient evidence to rate" may be noted, with no score assigned (i.e., left blank). Additional supplemental materials may even be solicited from the instructor or other sources at this stage, to help address unclear areas. After review of any additional sources of course information, the raters' individual assessments and notes should be compared, and then through discussion between the raters and additional consultation of all data sources available, an agreed-upon final rating for each element should be assigned. For this final scoring, no rating of "insufficient evidence to rate" should be included; instead, a score of zero (0) should be assigned for any element that is definitely absent or is still not evident after thorough review and discussion of the full set of available data sources. This procedure is in line with our guiding assumptions; because *every* element is considered important for service-learning quality, any element's absence intentionally and substantially reduces the course's overall summed quality score, as described next.

Establishing a Quality Score

To establish a total Service-Learning Quality Score for the course, the adjusted weighted ratings (which range from -25% to +25% of the base weights) for each of the 28 individual elements are summed. Because these elements have different base values representing their contribution to service-learning student outcomes, and these values are modified by level of implementation, the overall summed Quality Scores for any two given courses will typically vary. Relatedly, two courses may have the same overall Quality Score despite having different levels of presence, absence, and quality for particular elements.

A course scored as having all elements present at the baseline level thus receives a summed total Quality Score of 212.5. One in which all 28 elements are scored as present but all elements are below baseline would rate 159.5, and one in which all elements are present and above baseline would present a maximum possible score of 266.1. Our research has not yet established final guidelines for interpretation of these scores in

relation to other courses, nor where a cutoff point might be for a "high quality" course designation, for instance. However, the SLQAT provides a means to evaluate courses as having higher quality or lower quality in comparison to each other, allowing for more informed interpretations of the relationships between students' service-learning experiences and learning outcomes.

Discussion and Lessons Learned

In reflecting on our work over the past 6 years to create a reliable quantitative instrument to assess service-learning best practices, the complexity of this goal stands out. At the risk of stating the obvious, this is a difficult challenge. As our process description attests, deciding what is essential and what is not entails a judgment call informed by a large body of research and grounded practice. The question of what is universal in service-learning still seems open to potential differences in interpretation for different campuses and disciplines (e.g., Botelho et al., 2020), and becomes additionally complex when international contexts are considered. Even domestically, little evidence confirms whether enough consistency of practice exists between, for instance, first-year and graduate courses, or across different institutional types, or even among different groups of students, to allow use of a single, universal instrument. Different institutions may also place different emphasis on values embedded or explicit in their approach to service-learning, such as articulating social justice or critical service-learning, impacting judgments on what is essential in these courses.

Furthermore, gradations of quality are difficult to quantify and to describe, and even what seem like basic decisions (e.g., where to cut off between levels; how much spread is feasible in quantifying the implementation levels for each element) influence the form and use of the tool. Likewise, translating the essence of an element into descriptive language (describing what "baseline" implementation means, for instance) entails a balance between providing sufficient specificity to decide on a rating, without going too far in a particular direction that might limit application across diverse settings. Although our intent was to develop a quantitative instrument, a certain level of judgment, qualitative nuance, and individual variability seems likely to always remain inherent in holistically rating a course and its elements.

Other Recommendations for Practice

We originally conceptualized the SLQAT in order to develop quantitative, consistent overall quality scores allowing diverse service-learning courses to be rated in a more accurate and more nuanced way, in particular to allow for better institutional research on questions such as impact on student retention beyond the binary categorization of courses as "service-learning or not." We also envisioned this instrument as a key tool for a host of quantitative investigations, both as a predictor variable (e.g., "How well do higher SLQAT scores predict particular student outcomes?") and as a dependent variable (e.g., "What impact does faculty development programming have on course design and implementation?"). However, as was mentioned previously, the SLQAT also has the potential to impact practice and professional development beyond such research purposes. For instance, campuses and practitioners have expressed interest in using this tool for designing coursework, for reflective self-assessment of practice, and for ongoing quality improvement. Awareness of these key elements and their impacts could also support institutions in identifying what practices to include in their campus definitions and classifications of service-learning.

We suggest that the SLQAT can productively also serve as a basis for faculty development (or self-study) on the best practices of service-learning that promote positive student outcomes, and on key elements to consider when developing courses. As a self-assessment tool, the SLQAT can also provide practitioners with a quantitative score that indicates the level of overall quality (potentially benchmarked against other courses within and outside their institution) while also identifying particular elements of practice that are well implemented and those that may be improved. However, we also specifically advise against possible negative outcomes that could result from punitive adoption of an instrument such as this. Concerns have been raised that institutions or supervisors could attempt to use this tool to evaluate instructor teaching effectiveness. In our view, assessment of the quality of an instructor's teaching ability is not an appropriate use of the instrument, due to the complexity and contextualized variability of this pedagogical approach. In addition, the SLQAT focuses on the design and implementation of the service-learning

components of a course; it does not account for the nature, scope, or delivery of a course's academic content. Our research team also supports the idea that teaching and developing a service-learning course is an iterative and ongoing process that evolves with each implementation; SLQAT is designed to support instructors as they seek to implement the highest quality course that impacts student learning outcomes. Ideally, SLQAT would be used over time and provide positive support for instructors in this process of design, implementation, reflection, and redesign.

Limitations and Recommendations for Future Research

We acknowledge that the instrument and its development reflect premises that may not be universally accepted and have not yet been empirically assessed; however, these elements provide opportunities for future research more directly examining the decisions and assumptions of our research team as described in this article. In particular, we invite readers and researchers to consider the following caveats and areas for further investigation, and hope that the instrument will provide the impetus and opportunity to test (and ultimately support, disprove, or extend) these tenets. Likewise, we anticipate that the larger scale piloting process described below will also further validate some of the premises related to the instrument's development and use.

First, careful attention should be paid to how the elements were selected and operationalized in the SLQAT development process. Although the instrument is grounded in both research literature and the experience and expertise of those who helped shape, review, and pilot it, we acknowledge that the essentialness of each element has not been fully tested and should be evaluated further through additional research. Since the raters who provided the current baseline weights were not viewing the full version of the SLQAT instrument and approached service-learning work through different lenses and sets of experiences, we cannot ascertain whether they were interpreting these elements in the same way. Additionally, the SLQAT intentionally does not take into account a host of exogenous variables that likely influence the delivery of the course, such as instructors' experience, community or societal circumstances (e.g., a global pandemic), or unexpected circumstances such as changes in community partnership arrangements or staff during a semester or course offering. The elements included in the tool are only those over which the instructor has control.

We also note that the SLQAT is based on norms of practice and service-learning literature situated in Western and Northern education systems and practices. The intent of its development was to create a broadly applicable instrument, and international scholars were part of the pilot rating and feedback process; however, we do not yet have sufficient pilot testing with international courses to assert whether additional adaptation may be necessary for non-U.S. contexts. Although the development of the instrument was guided by assumptions related to universality of application in higher education contexts, we encourage practitioners and researchers to further test the breadth of that applicability in practice. In addition, given that the components that comprise the SLQAT are considered essential elements of service-learning, we also encourage further testing of the instrument within K–12 education contexts to assess the tool's applicability and utility in assessing quality service-learning experiences in primary and secondary school settings.

Future research should also more directly assess the assumption that higher SLQAT scores (i.e., "better" courses) bring about better student outcomes. As described earlier, the focus of the SLQAT and the selection of elements was intentionally oriented exclusively toward *student learning* outcomes. This focus, of course, does not capture the full importance and value of service-learning experiences; thus, the SLQAT likely excludes elements that impact or provide value to the community, instructor, or institution, to name some other possible stakeholders. The instrument also does not attempt to differentiate across the different types of student-level outcomes of interest to our field (e.g., academic learning, civic learning, graduation/retention, social–emotional, or character development). However, further research may productively investigate the relationship between the summed SLQAT Quality Score and any, all, or some of these student outcomes. Similarly, pilot participants have wondered whether single elements, or even composite dimension subscores, may have a standalone value as predictors of student outcomes, or whether the overall summed Quality Score is indeed the best metric.

Future research may thus help clarify the strength of the relationships between individual and collective elements of the rubric and particular student outcomes.

Though we treat the SLQAT's 28 elements as discrete, independent best practices in course design and implementation, relationships that influence the ways they are ultimately applied likely exist between and among them. For example, better "use of resources and support" (Element #9) might result in better "articulation of service-learning in syllabus" (Element #1) and/or more student reflection (Element #2); courses that clearly identify an "authentic community-based need" (Element #16) may likewise better demonstrate "mutual benefit" (Element #6), and so on. We also acknowledge that the current baseline weights, although informed by expert ratings, are still somewhat arbitrary; thus, there may or may not be a meaningful difference in impact between (for example) elements weighted with a 7.0 and a 7.5. Likewise, we hypothesize that a sum Quality Score for a course lacking some elements can validly be compared with that of a course that has all elements present; however, we have not yet tested this assumption.

Next Steps

Additional assessment of the SLQAT is needed to more fully validate the instrument as an accurate and effective measure of service-learning course quality. The research team is currently soliciting course materials (both foundational and supplemental) to be used for next-stage pilot testing of the instrument with an intentionally diverse set of courses. Ideally, this corpus of materials will represent service-learning courses modeling diverse approaches and settings (direct service, indirect service, graduate courses, undergraduate courses, first year seminars, etc.), different fields/departments, different institutional types and locations, and different levels of course quality (i.e., not just exemplary courses).

The next phase of piloting planned involves recruiting, training, and organizing a group of reviewers to evaluate course materials using the SLQAT and to ascertain reliability. We envision bringing together—virtually or in person—a set of raters to participate in training with the research team, then to rate, discuss, and debrief multiple courses, following the scoring guidelines and protocols described above. In addition to determining traditional measures of interrater reliability, other aspects of the SLQAT's validity and usability will be further investigated via rater feedback and reflection regarding time needed, challenges, and concerns about wording or operationalization. This piloting experience will help develop and inform content for future rater training, including confirmed, consistent element ratings for sample courses, explanations or definitions of terms used, and guidance regarding how evidence is used to achieve these ratings.

Additionally, the research team is collaborating with GivePulse to develop an online version of the instrument in order to facilitate its use and interpretation of results. This platform would automatically calculate summed Quality Scores and subscores as well as provide enhanced data displays to facilitate cross-rater comparisons. We further envision access to detailed scoring guidelines and training, comparative outcomes from multiple courses, and other online tools supporting the use of the SLQAT for both professional development and research purposes.

Conclusions

We set out to develop an instrument to meet an identified need for quantitative, more standardized rating of the key aspects of effective service-learning courses. Despite an investment of over 6 years, this result is in some ways a still unfinished attempt to quantify the quality of service-learning, a task that has proven much more complex than anticipated. We realize this is not necessarily the final version of the tool, which may be modified as we learn more from research in the field and as new dimensions of service-learning practice emerge. The instrument is complex by design in its structure and content, and requires time and practice to develop understanding of its various components and how best to use it. The effort and process of conceptualizing and building this instrument, reflecting upon the elements and descriptors, and considering the nuances and challenges of implementation, have been a worthwhile and rewarding experience for our research team. An instrument such as the SLQAT represents a valuable potential addition to research and practice for our field, and we invite other researchers and practitioners to use it as a starting point on their campuses

and beyond, and to evaluate and use it to better contribute to research, piloting, and reflective dialogue.

Acknowledgments

The contents of this article were developed in part under grant #P116140033 from Fund for the Improvement of Postsecondary Education, First in the World program, the U.S. Department of Education. However, the contents do not necessarily represent the policy of the U.S. Department of Education, and endorsement by the Federal Government should not be assumed.

About the Authors

Paul H. Matthews *is a senior academic professional and associate director of the University of Georgia Office of Service-Learning.*

Isabel Lopez *is the external research manager for Boston Public Schools.*

Laurel E. Hirt *is the director of the Center for Community-Engaged Learning at the University of Minnesota, Twin Cities.*

Shannon O. Brooks *is a senior academic professional and director of the University of Georgia Office of Service-Learning.*

Andrew Furco *is a professor in organizational leadership, policy, and development and director of the International Center for Research on Community Engagement at the University of Minnesota.*

References

Bailis, L., & Melchior, A. (2003). Practical issues in the conduct of large-scale, multisite research and evaluation. In S. H. Billig & A. S. Waterman (Eds.), *Studying service-learning: Innovations in education research methodology* (pp. 125–147). Routledge.

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.), *Creating our identities in service-learning and community engagement* (pp. 131–157). Information Age Publishing.

Billig, S. H., Root, S., & Jesse, D. (2005). The relationship between the quality indicators of service-learning and student outcomes: Testing professional wisdom. In S. Root, J. Callahan, & S. H. Billig (Eds.), *Improving service-learning practice: Research on models to enhance impacts* (pp. 97–115). Information Age Publishing.

Botelho, J., Eddy, R. M., Galport, N., & Avila-Linn, C. (2020). Uncovering the quality of STEM service-learning course implementation and essential elements across the California State University system. *Michigan Journal of Community Service Learning, 26*(2). https://doi.org/10.3998/mjcsloa.3239521.0026.201

Brandenburger, J. (2013). Investigating personal development outcomes in service learning: Theory and research. In P. H. Clayton, R. G. Bringle, & J. A. Hatcher (Eds.), *Research on service learning: Conceptual frameworks and assessment, Vol. 2A* (pp. 133–156). Stylus.

Bringle, R. G., Clayton, P. H., & Hatcher, J. A. (2013). Research on service learning: An introduction. In P. H. Clayton, R. G. Bringle, & J. A. Hatcher (Eds.), *Research on service learning: Conceptual frameworks and assessment, Vol. 2A* (pp. 3–25). Stylus.

Bringle, R. G., & Hatcher, J. A. (1995). A service-learning curriculum for faculty. *Michigan Journal of Community Service Learning, 2*(1), 112–122. http://hdl.handle.net/2027/spo.3239521.0002.111

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. http://hdl.handle.net/2027/spo.3239521.0016.203

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

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*(4), 233–245. https://doi.org/10.1080/00986280903172969

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

Furco, A. (2003). Issues of definition and program diversity in the study of service-learning. In S. H. Billig & A. S. Waterman (Eds.), *Studying service-learning: Innovations in education research methodology* (pp. 13–34). Routledge.

Furco, A., Brooks, S. O., Lopez, I., Matthews, P. H., Hirt, L. E., Schultzetenberg, A., & Anderson, B. N. (2023). Service-Learning Quality Assessment Tool (SLQAT). *Journal of Higher Education Outreach and Engagement, 27*(2), 181–200.

Hahn, T. W., Hatcher, J. A., Price, M. F., & Studer, M. L. (2016). *IUPUI taxonomy for service learning courses—Course design centric for institutional assessment and research*. http://hdl.handle.net/1805/10851

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. http://hdl.handle.net/2027/spo.3239521.0011.104

Heffernan, K. (2001). *Fundamentals of service-learning course construction*. Campus Compact.

Jacoby, B. (2015). *Service-learning essentials: Questions, answers, and lessons learned*. Jossey-Bass.

Jenkins, A., & Sheehey, P. (2011). A checklist for implementing service-learning in higher

education. *Journal of Community Engagement and Scholarship*, *4*(2), 52–60. https://doi.org/10.54656/XKNT9046

Kieran, L., & Haack, S. (2018). PRELOAD: A rubric to evaluate course syllabi for quality indicators of community engagement and service-learning components. *Journal of Community Engagement and Higher Education*, *10*(2), 39–47. https://discovery.indstate.edu/jcehe/index.php/joce/article/view/460

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice-Hall.

Kuh, G. D. (2008). Excerpt from "High-impact educational practices: What they are, who has access to them, and why they matter." *Association of American Colleges and Universities*, *14*(3), 28–29.

Langhout, R. D., Lopezzi, M. A., & Wang, Y.-C. (2023). Not all service is the same: How service-learning typologies relate to student outcomes at an Hispanic-serving institution. *Journal of Higher Education Outreach and Engagement*, *27*(2), 73–90.

Lockeman, K. S., & Pelco, L. E. (2013). The relationship between service-learning and degree completion. *Michigan Journal of Community Service Learning*, *20*(1), 18–30. http://hdl.handle.net/2027/spo.3239521.0020.102

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*(1), 32–47. http://hdl.handle.net/2027/spo.3239521.0005.104

Matthews, P. H., Dorfman, J. H., & Wu, X. (2015). The impacts of undergraduate service-learning on post-graduation employment outcomes. *International Journal of Research on Service-Learning and Community Engagement*, *3*(1). https://journals.sfu.ca/iarslce/iarslce/index.php/journal/article/view/109

Mungo, M. H. (2017). Closing the gap: Can service-learning enhance retention, graduation, and GPAs of students of color? *Michigan Journal of Community Service Learning*, *23*(2), 42–52. https://doi.org/10.3998/mjcsloa.3239521.0023.203

Provencher, A., & Kassel, R. (2017). High-impact practices and sophomore retention: Examining the effects of selection bias. *Journal of College Student Retention: Research, Theory & Practice*, *21*(2). https://doi.org/10.1177/1521025117697728

Roldan, M., Strage, A., & David, D. (2004). A framework for assessing academic service-learning across disciplines. In M. Welch & S. H. Billig (Eds.), *New perspectives in service-learning: Research to advance the field* (pp. 39–59). Information Age Publishing.

Shumer, R. (2003). Self-assessment for service-learning. In S. H. Billig & A. S. Waterman (Eds.), *Studying service-learning: Innovations in education research methodology* (pp. 149–171). Routledge.

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

Steinke, P., Fitch, P., Johnson, C., & Waldstein, F. (2002). An interdisciplinary study of service-learning predictors and outcomes among college students. In S. H. Billig & A. Furco (Eds.), *Service-learning through a multidisciplinary lens* (pp. 73–102). Information Age Publishing.

Stokamer, S. T. (2018). The intersection of institutional contexts and faculty development in service-learning and community engagement. In B. Berkey, C. Meixner, P. M. Green, & E. A. Eddins (Eds.), *Reconceptualizing faculty development in service-learning/community engagement: Exploring intersections, frameworks, and models of practice* (pp. 221–239). Stylus Publishing.

Warren, J. L. (2012). Does service-learning increase student learning? A meta-analysis. *Michigan Journal of Community Service Learning*, *18*(2), 56–61. http://hdl.handle.net/2027/spo.3239521.0018.205

Waterman, A. S. (2003). Issues regarding the selection of variables for study in the context of the diversity of possible student outcomes of service-learning. In S. H. Billig

& A. S. Waterman (Eds.), *Studying service-learning: Innovations in education research methodology* (pp. 73–90). Erlbaum.

Welch, M. (2010). O.P.E.R.A.: A first letter mnemonic and rubric for conceptualising and implementing service learning. *Issues in Educational Research, 20*(1), 76–82. http://www.iier.org.au/iier20/welch.html

Wilder, S. O., Berle, D., Knauft, D., & Brackmann, S. (2013). Long-term effects of service-learning on students' civic engagement after college. *Journal of Community Engagement and Scholarship, 6*(1), 125–126.

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

Appendix

The Service-Learning Quality Assessment Tool (SLQAT)—Dimension, Element Titles, and Short Descriptions

For full version of instrument see: Furco, A., Brooks, S. O., Lopez, I., Matthews, P. H., Hirt, L. E., Schultzetenberg, A., & Anderson, B. N. (2023). Service-Learning Quality Assessment Tool (SLQAT). *Journal of Higher Education Outreach and Engagement*, 27(2), 183–202.

Dimension I: Course Design Dimension (10 Elements)

Element #1: Articulation of Service-Learning in Syllabus
Service-learning is articulated and integrated in the course design and syllabus.

Element #2: Reflection
The course includes relevant critical reflection activities intended to foster connections between course content and service activities.

Element #3: Diverse Perspectives
The course provides opportunities to explore diverse perspectives on issues connected to goals/objectives and service activities.

Element #4: Assessment of Student Performance
The course incorporates assessment of students' performance related to service-learning experience.

Element #5: Flexibility in Course Design/Implementation
The course shows flexibility to evolve and adapt to community and student circumstances.

Element #6: Mutual Benefit
The service-learning experience is designed to benefit all stakeholders involved.

Element #7: Feedback
Stakeholders are given opportunities to provide feedback on the strengths and weaknesses of service-learning activities, design, and practices.

Element #8: Risk Management
Consideration of risk management is relevant and appropriate for the course and service activities.

Element #9: Use of Resources and Support for Service-Learning
The course makes use of available institutional or external supports for service-learning.

Element #10: Planning and Articulation of Service Activity
Details and specific expectations for the service activities are planned and articulated.

Dimension II: Learning Dimension (7 Elements)

Element #11: Academic Content Learning from Service-Learning
The service-learning experience's relationship to the academic content of the course is explicit, transparent, and rigorous.

Element #12: Societal Issues Learning from Service-Learning
The service-learning experience engages students in learning about societal issue[s] in explicit, transparent, relevant ways.

Element #13: Personal or Professional Learning from Service-Learning
The service-learning experience engages students in developing personal learning and/or professional skills.

Element #14: Appropriateness of Service Activities for Students
The service activities are contextually appropriate for students' level of skill/knowledge/experience.

Element #15: Connection between Service and Learning
The service activities and learning goals/objectives are linked.

Element #16: Authentic Community-Based Need
The service activities are based on a clear, meaningful community-identified issue/need.

Element #17: Appropriate Duration/Intensity of Service
The service activity's duration or intensity seems appropriate for community needs and course learning goals.

Dimension III: Student Dimension (3 Elements)

Element #18: Student Preparedness for Service-Learning
Students are prepared for the service-learning experience.

Element #19: Relevance of Service Activity
The course helps clarify the service-learning experience's relevance to students' interests, lives, etc.

Element #20: Student Voice
The course incorporates opportunities/activities for student voice (e.g., autonomy, choice, creativity, leadership, influence) in the service-learning experience.

Dimension IV: Instructor Dimension (3 Elements)

Element #21: Instructor's Knowledge of Service-Learning Pedagogy
The instructor has knowledge about service-learning pedagogy and expertise in how to apply it.

Element #22: Instructor's Knowledge of Community
The instructor is knowledgeable about community partners, contexts, needs, and norms.

Element #23: Instructor's Knowledge of Societal Issues
The instructor has understanding of the societal issue(s) that undergird the service-learning experience.

Dimension V: Community Partner and Partnership Dimension (5 Elements)

Element #24: Site/Partner Appropriateness
Service partners or locations are appropriate, given focus of course, level of students, focus of societal issue.

Element #25: Guidance and Supervision of Students
The community partner provides supervision, training, direction, and/or guidance to support students' experience.

Element #26: Community Partner Co-Educator Role
Community partners have a co-educator role and provide input in shaping the service-learning experience.

Element #27: Community Capacity for Service-Learning
Community partners have the capacity to support and participate fully in the service-learning experience.

Element #28: Instructor and Community Partner Connection
A partnership or relationship exists between the instructor and the community or community partner(s) for service-learning.