



# Test-Retest Reliability and Validity of the Service-Learning Outcomes Measurement Scale

**Ka Hing Lau**

*Lingnan University, Hong Kong*

**Robin S. Snell**

*Lingnan University, Hong Kong*

## Recommended Citation:

Lau, K. H., & Snell, R. S. (2020). Test-retest reliability and validity of the service-learning outcomes measurement scale. *International Journal of Research on Service-Learning and Community Engagement*, 8(1), Article 7. <https://doi.org/10.37333/001c.18782>

## Test-Retest Reliability and Validity of the Service-Learning Outcomes Measurement Scale

**Ka Hing Lau**

*Lingnan University, Hong Kong*

**Robin S. Snell**

*Lingnan University, Hong Kong*

The current study evaluated the test-retest reliability and validity of the 56-item *Service-Learning Outcomes Measurement Scale (S-LOMS)* by inviting 122 university students to respond to items on the scale twice within a two-week period. Results indicated test-retest reliability that was moderate for eight out of 11 domains and good for the three other domains. There were intraclass correlation coefficients (ICCs) of 0.6 or above at the domain, category, and overall levels of the scale. Notably, respondents with prior service-learning experience tended to score higher on developmental outcome domains than those without prior service-learning experience. Comparisons between *S-LOMS* scores according to gender, age, year of study, and academic background were also made and discussed. The results indicated that *S-LOMS* is stable over time and that it has robust validity vis-à-vis undergraduate students in Hong Kong as the target population.

**Keywords:** *Hong Kong, scale validity, service-learning, student learning outcomes, test-retest reliability*

### **Cuestionario de buena fiabilidad (test-retest) y validez de la Escala de Medición de los Resultados del Aprendizaje en Servicio**

En el presente estudio se evaluó la buena fiabilidad (procedimiento test-retest) y la validez de los 56 artículos incluidos en la Escala de Medición de Resultados del Aprendizaje en Servicio (S-LOMS según sus siglas en inglés). 122 estudiantes universitarios tomaron la encuesta en dos oportunidades en un período de dos semanas. Los resultados indicaron que la fiabilidad del test-retest era moderada en ocho de las 11 áreas y buena en las otras tres. Los coeficientes de correlación intraclase (ICC) obtenidos fueron de 0,6 o más en el dominio, categoría y niveles generales de la escala. Cabe destacar que los encuestados con experiencia previa de aprendizaje en servicio tendieron a obtener una puntuación más alta en el área de resultados de desarrollo que los que no tenían ninguna. También se establecieron y discutieron comparaciones entre las puntuaciones del S-LOMS según género, edad, año de estudio y antecedentes académicos. Los resultados indicaron que el S-LOMS es estable a lo largo del tiempo y que tiene una sólida validez con respecto a los estudiantes universitarios de pregrado en Hong Kong, población seleccionada.

**Palabras clave:** *Hong Kong, fiabilidad, validación de la escala, resultados del aprendizaje en servicio, cuestionario de buena fiabilidad (test-retest)*

**Editors' Note:** Translation by **Yamilet Hernandez**  
Department of English and Foreign Languages  
Barry University, USA

Service-learning was first adopted in Hong Kong tertiary education during the first decade of the 21st century (Ma, 2018). Service-learning is defined as “a form of experiential education in which students engage in activities that address human and community needs together with structured opportunities intentionally designed to promote student learning and development” (Jacoby 1996, p. 5). The introduction of service-learning to Hong Kong has taken place in the context of broader educational reforms, such as extending the undergraduate degree program from three to four years, and these reforms

have placed greater emphasis on whole-person development and on cultivating students to become socially responsible citizens (Freake, 2012; Ma, 2018; Shek et al., 2015).

Such reforms have thus encouraged tertiary institutions in Hong Kong to adopt service-learning (Xing & Ma, 2010; Snell & Lau, 2020). All eight government-funded universities there have established service-learning programs, along with at least one private local university, The Hang Seng University of Hong Kong. Lingnan University and the Hong Kong Polytechnic University have even made service-learning a graduation requirement (see Ma, 2018; Snell & Lau, 2020). An implication of this heightened commitment to service-learning by Hong Kong institutions is that assessing the effectiveness of this pedagogy in fostering the desired student developmental outcomes has become a shared priority. However, there has hitherto been a paucity of relevant research (Xing & Ma, 2010), due to the absence of a widely supported, standardized, and reliable measurement instrument (Snell & Lau, 2020).

To address this research gap, a 56-item *Service-Learning Outcomes Measurement Scale (S-LOMS)* was developed by Snell & Lau (2020) as a means to capture commonly desired developmental outcomes for students arising from service-learning. *S-LOMS* measures developmental outcomes in 11 domains under four major categories. The first category is knowledge application, with a single domain of the same name. The second category is personal and professional skills, comprising the domains of creative problem-solving skills, relationship and team skills, self-reflection skills, and critical thinking skills. The third category is civic orientation and engagement, comprising the domains of community commitment and understanding, caring and respect, and sense of social responsibility. The fourth category is self-awareness, comprising the domains of self-efficacy, self-understanding, and commitment to self-improvement. While these categories and domains are not dissimilar to those identified by service-learning scholars internationally (see Eyler et al., 2001; Jacoby, 1996; Snell & Lau, 2020), *S-LOMS* carries some unique merits. First, the sub-scales included in *S-LOMS* directly address developmental domains that local practitioners of service-learning across the higher educational sector regard as tailored to institutional values in Hong Kong. Second, *S-LOMS* brings these desired student developmental outcome domains together under a single, standardized scale. Third, *S-LOMS* offers flexibility for instructors or institutions to select particular domains for measurement, according to the priority outcomes addressed by their respective curricula (see Snell & Lau, 2020). Fourth, it follows that the availability of *S-LOMS* may facilitate the growth of research on the developmental impacts of service-learning on students in Hong Kong and potentially in other Asian jurisdictions.

The validity and reliability of *S-LOMS* has already undergone testing in earlier studies at four local universities in Hong Kong. A 102-item draft version of *S-LOMS* was initially piloted with local university students, for the purpose of improving the wording of items, before being administered to 400 students, whose responses were subjected to exploratory factor analysis (EFA). The EFA enabled the draft scale to be reduced to 56 items under the 11 domains, with all items having a factor loading of at least 0.40 on their respective domains, with reliability ranging between 0.75 and 0.96, and with variance explained for the four categories ranging from 67.7% to 72.4% (Snell & Lau, 2020).

In another prior study, Lau and Snell (2021) employed confirmatory factor analysis (CFA) on the responses of a fresh sample of 629 local university students to validate the factor structure that had been obtained in the EFA. The CFA confirmed the factor structure of 11 domains, with all items having a factor loading 0.5 or above. There were satisfactory goodness of fitness indices for the overall model (CFI: 0.91, NNFI: 0.90, RMSEA: 0.05), along with reliability scores of between 0.79 and 0.92 for the various domains. This factor structure was stable across demographic subgroups, such as men and women.

The current study sought to provide further validation of *S-LOMS* with respect to its test-retest reliability at the domain and category levels. Test-retest reliability aims “to assess the degree to which the participant’s performance is repeatable, i.e., how consistent their scores are across time” (Boateng et al., 2018, p. 13). Unlike the pretest-posttest approach that is normally employed for the purpose of assessing pedagogical effectiveness by collecting responses before and after the service-learning experience, test-retest reliability indicates whether the scale is stable across time when answered by the same respondents in the absence of the expected intervention.

In this study, the expected intervention was a service-learning experience. *S-LOMS* is developed to be administered on a pretest-posttest basis with service-learning as an intervention in between. In this way, the developmental outcomes created by service-learning can be reflected via the pretest-posttest score differences across the developmental domains and categories. Following this logic, the current study tested the scale validity by comparing changes in scores on *S-LOMS* over a relatively short time interval, during which it was anticipated that most students would not concurrently be engaged in a substantial service-learning experience. The current study also compared the scores of different subgroups in the sample. It was expected that the *S-LOMS* scores would not differ significantly between the beginning and end of a short time interval. In addition, it was expected that those students who had undertaken service-learning prior to the study would obtain significantly higher *S-LOMS* scores than those who had not done so. Moreover, score comparisons by gender, age, year of study, and academic background were also examined.

## Method

### Instrument

As explained above, *S-LOMS* has been designed to collect self-report data about students' development in 11 domains under four broader categories. Each of the domains comprises three to four items written in English, which are presented as statements describing different attributes of the respondent. Respondents are asked to indicate the extent to which they agree or disagree with each item on a 10-point Likert scale, from 1, strongly disagree, to 10, strongly agree (see Appendix 1 for the *S-LOMS* questionnaire).

### Procedure

The data gathering comprised two phases for each of the cohorts of students in the study. In the first (test) phase, students from three local universities, namely Lingnan University, The Hong Kong Polytechnic University, and The Education University of Hong Kong, were invited to complete *S-LOMS* and to provide information about their gender, age, year of study, academic background, and prior and current service-learning experiences, under assurance of data confidentiality. Explanations about our definition of service-learning and about how to categorize one's academic background were given to the students on request. Altogether, 266 students completed the instrument in the test phase. For each cohort, the students were invited to complete the same scale two weeks later. After this second (retest) phase, we were able to match test and retest responses by means of the respondents' student numbers. By the end of the data collection period, 122 students had finished both test and retest phases, with a completion rate of 45.9%. Each respondent was offered a gift voucher as a token for their participation upon completing both test and retest phases.

It was preferred that respondents would not be undertaking any service-learning programs during the time interval between the test and retest phases. With this purpose in mind, data were collected between May and October 2019, which covers the summer vacation and the beginning of the fall semester. The test-retest period for each student was set at two weeks. We considered that this interval was sufficiently long to reduce any carryover effect arising from respondents' ability to recall their initial responses to the items, while also being sufficiently short to reduce the likelihood of a concurrent service-learning experience. An additional consideration was that an interval as short as two weeks would also reduce the impact on retest scores of any concurrent service-learning experience.

Data about whether respondents were undertaking any concurrent service-learning programs were collected. Since service-learning in Hong Kong normally requires at least one semester to complete, it was hypothesized that the test-retest differences between those who were concurrently taking service-learning programs and those who were not doing so would be statistically insignificant.

## Statistical Analysis

After the dataset was inputted and cleaned, test-retest reliability indices were derived for the scale at the domain, category, and overall levels. The reliability indices employed in this study comprised intraclass correlation coefficients (ICCs), based on a two-way mixed-effects model with absolute agreement, generated by using the software of SPSS version 25, as recommended by Koo and Li (2016). Since the measurement instrument is a single assessment tool that is designed to be self-rated by students before and after the service-learning experience, the reliability indices of single measures were accordingly adopted in the output tables (McGraw & Wong, 1996; Nichols, 1998). Koo and Li's (2016) formula for deriving the above ICCs was as follows:

$$ICC = MS_R - MS_E / [MS_R + (k-1) MS_E + (k/n) (MS_C - MS_E)]$$

where  $MS_R$  = mean square for rows;  $MS_E$  = mean square for error;  $MS_C$  = mean square for columns;  $n$  = number of respondents;  $k$  = number of raters/ measurements.

Although many studies have adopted Pearson's product-moment correlation for test-retest reliability, we employed ICCs because the former "merely assesses the extent to which scores go together and not whether they are close to each other in absolute terms" (Kazdin, 2011, p. 107). Furthermore, ICCs are able to capture the extent of homogeneity with absolute agreement. ICC scores were thus calculated for the entire sample of respondents.

In addition, scores at the domain, category, and overall levels of *S-LOMS* were compared between groups of respondents with different demographic characteristics, including gender, age, year of study, academic background, possession of prior service-learning experience, and whether they were concurrently undertaking service-learning programs. The scores for such intergroup comparisons were derived by taking the mean between the test and retest scores for each respondent. Reliability scores in terms of Cronbach's Alpha were also calculated.

## Results

### Demographics of the Respondents

Female respondents constituted a larger part of the sample (75.4%). The average age of the sample was 20.44 (SD = 1.86; range: 18–27) with 87.7% of the respondents between the ages of 18 and 22. The sample contained a plurality of students with an engineering and science background (35.2%), followed by those in the arts (34.4%), social sciences (13.9%), business (10.7%) and healthcare (5.7%) disciplines. There were 32 respondents from the 1st year (26.2%), 25 from the 2nd year (20.5%), 42 from the 3rd year (34.4%), 17 from the 4th year (13.9%), and 6 from the 5th year (4.9%). Among the respondents, 34.4% had prior service-learning experience while 25.4% were concurrently participating in service-learning programs.

### Test-Retest Reliability

Table 1 presents the ICC reliability indices and 95% confidence intervals, as well as the mean scores, for the 11 domains, for the four overarching categories, and for *S-LOMS* as a whole. The entire instrument obtained a reliability index of .78, while reliabilities ranged between .67 and .78 for the four overarching categories, and between .60 and .77 for the 11 constituent domains. According to the recommendation by Koo and Li (2016), most of the test-retest reliability index scores obtained from the current study are regarded as moderate (i.e., between .50 and .75) while the others, such as those for the category of civic orientation and engagement and for the domain of caring and respect, may be regarded as good (i.e., between .75 and .90). Cronbach's alpha values were obtained for the one-off reliability of the measurement instrument at both the testing and retesting times. These were satisfactory, comprising values between .74 and .93 at the domain level, between .86 and .96 at the category level, and at .98 for the overall scale.

**Table 1***The Intraclass Correlation Coefficients and Reliability Results (N = 122)*

Category/ Domain	Test-Retest (95% confidence interval)	ICC confidence	Cronbach's Alpha		Mean Score			
			Time 1	Time 2	Time 1		Time 2	
			$\alpha$	$\alpha$	$\bar{x}$	S.D.	$\bar{x}$	S.D.
<b>Knowledge Application</b>	<b>.67 (.56, .76)</b>		<b>.86</b>	<b>.90</b>	<b>6.90</b>	<b>1.14</b>	<b>7.04</b>	<b>1.14</b>
<b>Personal and Professional Skills</b>	<b>.73 (.63, .80)</b>		<b>.96</b>	<b>.96</b>	<b>7.05</b>	<b>1.06</b>	<b>7.17</b>	<b>1.04</b>
Creative Problem-Solving Skills	.72 (.62, .80)		.92	.93	6.98	1.17	7.20	1.19
Relationship and Team Skills	.74 (.64, .81)		.92	.92	6.87	1.26	6.91	1.22
Self-reflection Skills	.60 (.47, .70)		.80	.88	7.32	1.19	7.44	1.23
Critical Thinking Skills	.68 (.57, .77)		.79	.80	7.35	1.11	7.43	1.05
<b>Civic Orientation and Engagement</b>	<b>.78 (.70, .84)</b>		<b>.94</b>	<b>.95</b>	<b>7.31</b>	<b>1.06</b>	<b>7.41</b>	<b>1.04</b>
Community Commitment and Understanding	.75 (.66, .82)		.90	.91	6.95	1.11	7.09	1.09
Caring and Respect	.76 (.67, .83)		.92	.92	7.78	1.17	7.84	1.13
Sense of Social Responsibility	.70 (.59, .78)		.75	.80	7.18	1.35	7.23	1.32
Community Commitment and Understanding cum Sense of Social Responsibility	.77 (.68, .83)		.91	.92	7.01	1.09	7.13	1.08
<b>Self-awareness</b>	<b>.73 (.64, .80)</b>		<b>.91</b>	<b>.93</b>	<b>7.13</b>	<b>1.08</b>	<b>7.24</b>	<b>1.13</b>
Self-efficacy	.72 (.63, .80)		.84	.88	6.90	1.25	6.98	1.35
Self-understanding	.66 (.55, .75)		.83	.85	7.28	1.16	7.28	1.26
Commitment to Self-improvement	.62 (.48, .72)		.74	.82	7.23	1.23	7.52	1.16
<b>The Measurement Instrument</b>	<b>.78 (.70, .84)</b>		<b>.98</b>	<b>.98</b>	<b>7.14</b>	<b>1.01</b>	<b>7.24</b>	<b>0.98</b>

## Mean Scores Comparisons

The test and retest scores for each respondent on each item were averaged, and mean scores were calculated at the domain, category, and overall levels, for the purpose of making comparisons between groups. Statistical tests were then undertaken to compare mean scores between different demographic groups, with t-tests performed to investigate the impact of gender and year of study, and one-way ANOVAs employed to investigate the impact of academic background. Before conducting these ANOVAs, due to concerns about small numbers of respondents in some disciplines, and considering the relative proximity of the respective disciplines, the respondents were regrouped according to their major discipline. Those majoring in social sciences and business were put under one group; those from science, engineering, and healthcare were put under another group; those majoring in arts were kept as a separate group. Furthermore, the 1st- and 2nd-year respondents were put in one group and those from the 3rd year up were put in another group for performing the t-test because of the sample size concern. In addition, Pearson's correlations were obtained between respondents' age and scores.

Tables 2, 3, and 4 display the generally insignificant results found for demographic comparisons. Across gender, only the differences in mean scores for the domains of sense of social responsibility ( $p = .04$ ) and commitment to self-improvement ( $p = .02$ ) were significant. Respondents' ages, year of study, and academic background were not significantly associated with their *S-LOMS* scores.

**Table 2**

*Average Scores and T-test Results ( $df = 120$ ) Across Gender and Pearson's Correlation Between S-LOMS Scores and Age*

	Gender		Age				
	Male (N = 30)		Female (N = 92)		$p$	$\Delta\bar{x}$	$R$
Category/ Domain	$\bar{x}$	s.d.	$\bar{x}$	s.d.			
<b>Knowledge Application</b>	<b>6.80</b>	<b>1.21</b>	<b>7.02</b>	<b>.99</b>	<b>.32</b>	<b>.22</b>	<b>-.02</b>
<b>Personal and Professional Skills</b>	<b>6.92</b>	<b>1.21</b>	<b>7.17</b>	<b>.89</b>	<b>.30</b>	<b>.25</b>	<b>-.06</b>
Creative Problem-Solving Skills	6.88	1.29	7.16	1.03	.22	.28	.00
Relationship and Team Skills	6.70	1.47	6.95	1.03	.40	.25	-.11
Self-reflection Skills	7.15	1.31	7.46	.99	.18	.31	-.03
Critical Thinking Skills	7.32	1.10	7.42	.96	.63	.10	-.04
<b>Civic Orientation and Engagement</b>	<b>7.11</b>	<b>1.09</b>	<b>7.44</b>	<b>.94</b>	<b>.11</b>	<b>.33</b>	<b>-.02</b>
Community Commitment and Understanding	6.85	1.11	7.08	1.00	.31	.23	-.04
Caring and Respect	7.53	1.29	7.90	1.00	.11	.37	-.04
Sense of Social Responsibility	6.81	1.43	7.33	1.13	.04*	.52	.06
Community Commitment and Understanding cum Sense of Social Responsibility	6.84	1.11	7.15	.98	.16	.31	-.01
<b>Self-awareness</b>	<b>6.90</b>	<b>1.27</b>	<b>7.28</b>	<b>.93</b>	<b>.08</b>	<b>.38</b>	<b>.06</b>
Self-efficacy	6.66	1.37	7.03	1.14	.15	.37	.04
Self-understanding	7.07	1.48	7.35	.95	.33	.28	.11
Commitment to Self-improvement	6.98	1.19	7.51	1.01	.02*	.53	.00
<b>The Measurement Instrument</b>	<b>6.97</b>	<b>1.13</b>	<b>7.27</b>	<b>.86</b>	<b>.13</b>	<b>.30</b>	<b>-.02</b>

Note: \*:  $p \leq .05$

**Table 3**  
*Scores and One-Way ANOVA Results ( $df = 2, 119$ ) Across Academic Background*

Category/ Domain	a. Arts (N = 42)		b. Social Sciences & Business (N = 0)		c. Engineering, Science & Healthcare (N = 50)		<i>p</i>	$\Delta \bar{x}$		
	$\bar{x}$	s.d.	$\bar{x}$	s.d.	$\bar{x}$	s.d.		(a-b)	(a-c)	(b-c)
<b>Knowledge Application</b>	<b>7.09</b>	<b>1.15</b>	<b>7.07</b>	<b>.93</b>	<b>6.81</b>	<b>1.01</b>	<b>.38</b>	<b>.02</b>	<b>.28</b>	<b>.26</b>
<b>Personal and Professional Skills</b>	<b>7.15</b>	<b>1.06</b>	<b>7.22</b>	<b>.88</b>	<b>7.01</b>	<b>.97</b>	<b>.60</b>	<b>-.07</b>	<b>.14</b>	<b>.21</b>
Creative Problem-Solving Skills	7.10	1.17	7.17	1.06	7.04	1.07	.88	-.07	.06	.13
Relationship and Team Skills	6.94	1.21	7.05	1.14	6.75	1.12	.49	-.11	.19	.30
Self-reflection Skills	7.44	1.06	7.54	.95	7.23	1.17	.42	-.10	.21	.31
Critical Thinking Skills	7.47	1.08	7.39	1.01	7.32	.91	.78	.08	.15	.07
<b>Civic Orientation and Engagement</b>	<b>7.46</b>	<b>1.05</b>	<b>7.39</b>	<b>.92</b>	<b>7.26</b>	<b>.98</b>	<b>.61</b>	<b>.07</b>	<b>.20</b>	<b>.13</b>
Community Commitment and Understanding	7.12	1.05	7.06	.95	6.91	1.07	.61	.06	.21	.15
Caring and Respect	7.88	1.17	7.84	.98	7.74	1.08	.82	.04	.14	.10
Sense of Social Responsibility	7.38	1.13	7.21	1.12	7.05	1.36	.43	.17	.33	.16
Community Commitment and Understanding cum Sense of Social Responsibility	7.19	1.03	7.10	.94	6.95	1.07	.51	.09	.24	.15
<b>Self-awareness</b>	<b>7.26</b>	<b>1.05</b>	<b>7.38</b>	<b>.94</b>	<b>7.00</b>	<b>1.05</b>	<b>.22</b>	<b>-.12</b>	<b>.26</b>	<b>.38</b>
Self-efficacy	6.87	1.32	7.21	1.20	6.84	1.11	.36	-.34	.03	.37
Self-understanding	7.43	1.00	7.46	.98	7.05	1.22	.14	-.03	.38	.41
Commitment to Self-improvement	7.55	1.11	7.52	1.01	7.15	1.07	.14	.03	.40	.37
<b>The Measurement Instrument</b>	<b>7.27</b>	<b>1.03</b>	<b>7.30</b>	<b>.84</b>	<b>7.07</b>	<b>.93</b>	<b>.48</b>	<b>-.03</b>	<b>.20</b>	<b>.23</b>

Note: \*:  $p \leq .05$



**Table 4***Average Scores and T-test Results (df = 120) Across Year of Study*

Category/ Domain	Year of Study					
	Years 1 & 2 (N = 57)		Years 3, 4, & 5 (N = 65)			
	$\bar{x}$	s.d.	$\bar{x}$	s.d.	$p$	$\Delta \bar{x}$
Knowledge Application	6.90	1.11	7.03	.99	.48	.14
Personal and Professional Skills	7.07	1.02	7.14	.94	.69	.07
Creative Problem-Solving Skills	7.07	1.11	7.11	1.10	.85	.04
Relationship and Team Skills	6.83	1.26	6.94	1.06	.63	.10
Self-reflection Skills	7.34	1.15	7.41	1.03	.73	.07
Critical Thinking Skills	7.35	.97	7.43	1.01	.65	.08
Civic Orientation and Engagement	7.31	1.03	7.40	.95	.60	.10
Community Commitment and Understanding	7.00	1.10	7.04	.98	.82	.04
Caring and Respect	7.80	1.15	7.82	1.02	.90	.02
Sense of Social Responsibility	6.99	1.27	7.39	1.16	.07	.40
Community Commitment and Understanding cum Sense of Social Responsibility	7.00	1.07	7.14	.97	.45	.14
Self-awareness	7.07	1.09	7.28	.97	.25	.21
Self-efficacy	6.87	1.24	7.00	1.19	.54	.13
Self-understanding	7.08	1.17	7.45	1.01	.06	.37
Commitment to Self-improvement	7.32	1.17	7.43	1.00	.56	.11
The Measurement Instrument	7.14	.98	7.25	.90	.52	.11

*Note: \*:  $p \leq .05$*

**Table 5**

*Scores and T-test Results (df = 120) Between Groups (a) With and Without Prior Service-Learning Experience and (b) Whether Currently Taking Service-Learning Programs*

Category/ Domain	Possessing Experience		Prior		Service-Learning		Whether Currently Taking Service-Learning Programs or Courses					
	Yes (N = 42)		No (N = 80)				Yes (N = 31)		No (N = 91)			
	$\bar{x}$	s.d.	$\bar{x}$	s.d.	<i>p</i>	$\Delta\bar{x}$	$\bar{x}$	s.d.	$\bar{x}$	s.d.	<i>p</i>	$\Delta\bar{x}$
<b>Knowledge Application</b>	<b>7.10</b>	<b>.99</b>	<b>6.90</b>	<b>1.07</b>	<b>.33</b>	<b>.20</b>	<b>7.05</b>	<b>1.14</b>	<b>6.94</b>	<b>1.02</b>	<b>0.61</b>	<b>.11</b>
<b>Personal and Professional Skills</b>	<b>7.31</b>	<b>.88</b>	<b>7.00</b>	<b>1.01</b>	<b>.10</b>	<b>.31</b>	<b>7.26</b>	<b>1.11</b>	<b>7.06</b>	<b>0.93</b>	<b>0.32</b>	<b>.20</b>
Creative Problem-Solving Skills	7.20	1.07	7.04	1.11	.45	.16	7.28	1.20	7.03	1.06	0.28	.25
Relationship and Team Skills	7.20	.93	6.72	1.23	.03*	.48	7.10	1.18	6.82	1.14	0.23	.28
Self-reflection Skills	7.59	.93	7.27	1.14	.12	.32	7.46	1.25	7.35	1.03	0.65	.11
Critical Thinking Skills	7.55	.87	7.31	1.04	.17	.24	7.39	1.02	7.39	0.98	1.00	.00
<b>Civic Orientation and Engagement</b>	<b>7.54</b>	<b>.89</b>	<b>7.26</b>	<b>1.03</b>	<b>.14</b>	<b>.28</b>	<b>7.46</b>	<b>1.00</b>	<b>7.33</b>	<b>0.99</b>	<b>0.51</b>	<b>.13</b>
Community Commitment and Understanding	7.19	.93	6.93	1.08	.19	.26	7.21	1.03	6.96	1.03	0.24	.25
Caring and Respect	7.97	.95	7.73	1.14	.21	.24	7.80	1.08	7.81	1.09	0.95	-.01
Sense of Social Responsibility	7.47	1.13	7.06	1.26	.08	.41	7.33	1.19	7.16	1.24	0.50	.17
Community Commitment and Understanding, and Sense of Social Responsibility	7.27	.93	6.97	1.06	.12	.30	7.24	1.00	7.01	1.03	0.28	.23
<b>Self-awareness</b>	<b>7.50</b>	<b>.95</b>	<b>7.01</b>	<b>1.04</b>	<b>.01**</b>	<b>.49</b>	<b>7.20</b>	<b>1.17</b>	<b>7.18</b>	<b>0.98</b>	<b>0.90</b>	<b>.02</b>
Self-efficacy	7.33	1.05	6.73	1.24	.01**	.60	6.96	1.27	6.93	1.19	0.90	.03
Self-understanding	7.64	1.10	7.09	1.06	.01**	.55	7.34	1.30	7.26	1.03	0.74	.08
Commitment to Self-improvement	7.55	0.94	7.29	1.14	.20	.26	7.34	1.27	7.39	1.02	0.84	-.05
<b>The Measurement Instrument</b>	<b>7.41</b>	<b>.85</b>	<b>7.08</b>	<b>0.97</b>	<b>.07</b>	<b>.33</b>	<b>7.30</b>	<b>1.05</b>	<b>7.16</b>	<b>0.90</b>	<b>0.47</b>	<b>.14</b>

Notes: \*:  $p \leq .05$ ; \*\*:  $p \leq .01$

Another set of statistical tests compared the mean scores between students with prior service-learning experience and those without such prior experience, and between students who were currently taking part in service-learning programs versus those who were not concurrently taking part in any. Table 5 reports the respective mean scores and t-test results.

Results indicated that the respondents with prior service-learning experience scored significantly higher than those without prior service-learning experience on three domains, i.e., relationship and team skills ( $p = .03$ ), self-efficacy ( $p = .01$ ), and self-understanding ( $p = .01$ ), and on one category, i.e., self-awareness ( $p = .01$ ). However, the respondents who were concurrently participating in service-learning programs were not found to score significantly higher on any domains as compared with those who were not concurrently participating in service-learning programs.

Moreover, the mean score differences between the respondents with prior service-learning experience and those without such prior experience were generally higher than the mean score differences between the respondents currently taking service-learning programs and those who were not doing so. This contrast is consistent with previous research findings about the positive relationship between the duration of the service-learning experience and the ensuing beneficial impacts on the development of participating students (Astin & Sax, 1998).

## Discussion

The current study established that, in terms of ICC, there was moderate test-retest reliability for eight of the 11 domains of *S-LOMS* and good test-retest reliability for the three other domains, and that there was also good test-retest reliability for the overall measurement instrument. Moreover, Cronbach's alpha results demonstrated that there was good scale and subscale reliability. At the category level, civic orientation and engagement obtained the highest ICC score, followed by self-awareness, tied with personal and professional skills, while knowledge application had the lowest ICC score. The higher ICC scores for civic orientation and engagement may reflect that this is a relatively stable trait over time and is not readily altered. By comparison, the lower ICC scores for knowledge application may reflect that some respondents, who may have been studying or receiving other forms of training during the two-week interval between test and retest, may also have recognized opportunities to apply knowledge that they had learned. Patterns of difference in ICC scores were also found at the domain level. Caring and respect, along with community commitment and understanding, obtained the highest ICC scores, whereas the domains of self-reflection skills and commitment to self-improvement appeared to be less stable.

*S-LOMS* was able to differentiate between those respondents with or without prior service-learning experience, and the mean score differences imply that the influence of prior completed service-learning experience can be substantial. Another relatively subtle yet expected result was that there were much smaller differences between the scores of respondents who were concurrently undertaking service-learning programs and those who were not doing so, reflecting the brevity of the test-retest interval of two weeks.

One possible explanation for the different developmental impacts of prior and concurrent service-learning experience is that such impacts take a certain amount of time to become manifest after initial stimulation. This explanation would be in line with findings of prior research on the long-term impact of service-learning on students' future community involvement (e.g., Astin et al., 2006), on students' self-awareness and volunteering behaviors (e.g., Reinders & Yourniss, 2009), and on students' identity formation and openness to new ideas (e.g., Jones & Abes, 2004).

Comparisons of *S-LOMS* scores across gender indicated that female respondents scored generally higher than their male counterparts, particularly in the domains of civic orientation and engagement and self-awareness; yet most of these differences were insignificant. For the domain of critical thinking skills, the difference was the smallest. This overall pattern of gender comparison results is not dissimilar to previous research findings, which indicate that females generally outperform males on criteria salient to educational contexts, except for reasoning ability (e.g., King et al., 1990; Leach & Good, 2011; Pomerantz et al., 2002).

That comparisons of *S-LOMS* scores across age yielded insignificant correlation coefficients probably reflects that nearly 90% of the respondents were between ages 18 and 22, thereby making the correlation analysis insufficiently sensitive. The finding that *S-LOMS* scores compared across academic background and year of study were insignificant may reflect both that the scale's domains are not discipline specific and that they are not so readily affected simply by progression through the undergraduate degree structure. For example, personal and professional skills, such as creative problem-solving and relationship and team skills, correspond to typical, university-wide desired graduate attributes, as do aspects of civic orientation and engagement, such as caring and respect and sense of social responsibility. Furthermore, developing those attributes at higher educational institutions of Hong Kong typically relies on extracurricular activities rather than on the regular syllabus provided by a student's respective major department, which is arranged according to discipline-based knowledge sequencing rather than on systematic attempts to foster student development in the learning domains that are encompassed by *S-LOMS*. More widespread and concerted adoption of service-learning across the universities may change this picture in future. All things considered, it is understandable that within the sample, *S-LOMS* scores did not differ significantly across students from different academic backgrounds and years of study. Overall, the comparisons of *S-LOMS* scores across respondents' demographics supported that *S-LOMS* is a reliable measurement instrument.

## Conclusions and Implications

The current study established that there is sound test-retest reliability for *S-LOMS*, in that *S-LOMS* displays sound patterns of reliability at both the category and domain levels, with eight out of 11 domains obtaining moderate ICCs and the three others achieving good ICCs.

The statistical findings about respondents' demographics provided further support for the validity of *S-LOMS*, as did the findings of differences between the *S-LOMS* scores of respondents with prior service-learning experience versus those without such experience. An implication of the latter findings is that completed service-learning experience has had positive developmental impacts for students on domains measured by *S-LOMS*.

By comparison, there were smaller differences in *S-LOMS* scores between those respondents who were concurrently undertaking service-learning programs versus those who were not concurrently undertaking any. This finding implies that an observation period that is longer than two weeks is required to detect the emergence of developmental outcomes in students resulting from the service-learning experience.

One limitation of the current study is that the potential impact of students' prior service-learning experience on their subsequent development through additional service-learning experiences has not been investigated. It is possible that prior service-learning experience serves as a moderator, positively affecting the developmental outcomes of students' subsequent service-learning experiences. Future research could examine longitudinally at various points in time the developmental impacts on students of undertaking a service-learning course or program after commencing it and after completing it. Studies could also use *S-LOMS* to analyze whether student growth in one or more developmental domains serves as a mediating or moderating factor for that student's growth in other salient developmental domains.

Advantages of the current study include that data were collected from three different government-funded universities and that there was some diversity in terms of the students' degree programs. Future studies should nonetheless seek to collect data from a broader range of institutions, including private universities, in order to generalize the validity and reliability of *S-LOMS* across the higher educational sector in Hong Kong, and perhaps extending into other Asian jurisdictions.

By providing evidence of sound test-retest reliability, this study builds on prior research into the *S-LOMS* dimensionality according to EFA and CFA (Lau & Snell, 2021; Snell & Lau, 2020). One further step in terms of scale validation would involve confirming the predictive validity of *S-LOMS* by using the scale to evaluate the developmental impacts of different service-learning programs, using a pretest-posttest design, and taking into account instructor expectations and priorities. In addition, *S-LOMS* scores

may be correlated with scores obtained via other scales for evaluating student developmental outcomes in similar domains.

## References

- Astin, A. W., & Sax, L. J. (1998). How undergraduates are affected by service participation. *Journal of College Student Development*, 39(3), 251–263.
- Astin, A. W., Vogelgesang, L. J., Misa, K., Anderson, J., Denson, N., Jayakumar, U., Saenz, V., & Yamamura, E. (2006). *Understanding the effects of service-learning: A study of students and faculty*. Higher Education Research Institute, University of California.  
[https://www.academia.edu/12347851/Understanding\\_the\\_effects\\_of\\_service\\_learning\\_A\\_study\\_of\\_students\\_and\\_faculty](https://www.academia.edu/12347851/Understanding_the_effects_of_service_learning_A_study_of_students_and_faculty)
- Boateng, G. O., Neilands T. B., Frongillo E., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 149(6), 1–18. <https://doi.org/10.3389/fpubh.2018.00149>
- Eyler, J. S., Giles, D. E., Stenson, C. M. & Gray, C. (2001). *At a glance: What we know about the effects of service-learning on college students, faculty, institutions and communities, 1993–2000* (3rd ed.). Vanderbilt University Press.
- Freake, H. (2012). Curricular designs for general education at the UGC-supported universities in Hong Kong. In J. Xing, P. S. Ng, & C. Y. Cheng (Eds.), *General education and the development of global citizenship in Hong Kong, Taiwan and Mainland China* (pp. 105–120). Routledge.
- Jacoby, B. (1996). *Service-learning in higher education: Concepts and practices*. Jossey-Bass.
- Jones, S. R., & Abes, E. S. (2004). Enduring influences of service-learning on college students' identity development. *Journal of College Student Development*, 45(2), 149–166.
- Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press.
- King, P., Wood, P., & Mines, R. (1990). Critical thinking among college and graduate students. *The Review of Higher Education*, 13(2), 167–186.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
- Lau, K. H., & Snell, R. S. (2021). Confirmatory factor analysis for a service-learning outcomes measurement scale (S-LOMS). *Metropolitan Universities*, 32(1), 3–34.
- Leach, B. T., & Good, D. W. (2011). Critical thinking skills as related to university students' gender and academic discipline. *International Journal of Humanities and Social Science*, 1(21), 100–106.
- Ma, C. (2018). Service-learning development in higher education in Hong Kong. In T. W. Lim & W. X. Li (Eds.), *Studying Hong Kong: 20 years of political, economic and social developments* (pp. 43–61). World Scientific. [https://doi.org/10.1142/9789813223554\\_0004](https://doi.org/10.1142/9789813223554_0004)
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30–46. <http://dx.doi.org/10.1037/1082-989X.1.1.30>
- Nichols, D. P. (1998). *Choosing an intraclass correlation coefficient*.  
<https://stats.idre.ucla.edu/spss/library/spss-library-choosing-an-intraclass-correlation-coefficient/>
- Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology*, 94(2), 396–404. <https://doi.apa.org/doi/10.1037/0022-0663.94.2.396>
- Reinders, H., & Yourniss, J. (2009). School-based required community service and civic development in adolescents. *Applied Developmental Science*, 10, 2–12. [https://doi.org/10.1207/s1532480xads1001\\_1](https://doi.org/10.1207/s1532480xads1001_1)
- Shek, D. T. L., Yu, L., Wu, F. K. Y., & Chai, W. Y. (2015). General university requirements at Hong Kong Polytechnic University: Evaluation findings based on student focus groups. *Assessment & Evaluation in Higher Education*, 40(8), 1017–1031. <https://doi.org/10.1080/02602938.2014.960362>

- Snell, R. S., & Lau, K. H. (2020). The development of a service-learning outcomes measurement scale (S-LOMS). *Metropolitan Universities*, 31(1), 44–77. <https://doi.org/10.18060/23258>
- Xing, J., & Ma, C. (Eds.). (2010). *Service-learning in Asia: Curricular models and practices*. Hong Kong University Press. <https://doi.org/10.5790/hongkong/9789888028467.001.0001>

## Appendix

### Items, Domains, and Categories of *S-LOMS*

Item No.	Domain	Item
Category: Knowledge Application		
1	Knowledge Application	I know how to transfer knowledge and skills from one setting to another.
2		I can make connections between theory and practice.
3		I am able to apply/integrate classroom knowledge to deal with complex issues.
4		I know how to apply what I learn in class to solve real-life problems.
Category: Personal and Professional Skills		
5	Relationship and Team Skills	I am good at building relationships between people.
6		I can easily establish effective relationships with people.
7		I can build long-term relationships with people.
8		I am good at keeping in touch with people.
9		I have the necessary skills for making groups or organizations function effectively.
10	Creative Problem-Solving Skills	I am good at resolving conflicts.
11		I am confident in leading others toward common goals.
12		I participate effectively in group discussions and activities.
13		I am able to solve challenging real-life problems.
14		I feel confident in dealing with a problem.
15		I often modify my strategies to solve a problem when the situation changes.
16		I feel confident in identifying the core of a problem.
17		I am not afraid of trying new things.
18		I am able to generate original ideas.
19		I am able to look at an issue from a fresh perspective.
20	Self-reflection Skills	When necessary, I can think of alternatives.
21		I always think how I can improve myself.
22		I will evaluate myself after completing a task.
23	Critical Thinking Skills	I consider circumstances when reflecting on how well I have performed.
24		I reflect on myself regularly.
25		I can analyze an issue comprehensively.
26		I often look at complex issues from different angles.
27		I can understand others' viewpoints when we are making decisions together.
Category: Civic Orientation and Engagement		
28	Community Commitment and Understanding	I think about how I can serve the community after graduating.
29		I will play my part to reduce social problems.
30		I always actively discuss possible improvements for our community.
31		I will contribute my abilities to make the community a better place.

32		I can identify challenges in the community.
33		I can investigate the challenges faced by people in need in a community.
34		I can identify issues that are important for a disadvantaged community.
35		I can identify useful resources of a community.
36	Caring and Respect	I can respect people whose background is different from mine.
37		I am willing to try to understand people whose background is different from mine.
38		I respect the needs of people from different backgrounds.
39		I appreciate the ideas of people from different backgrounds.
40		I observe others' feelings and emotions.
41		I consider others' points of view.
42		I care about others.
43	Sense of Social Responsibility	I believe that taking care of people who are in need is everyone's responsibility.
44		I feel obligated to help those who are less fortunate than me.
45		I believe that everybody should be encouraged to participate in civic affairs.
<hr/>		
<b>Category: Self-awareness</b>		
46	Self-efficacy	Most things I do, I do well.
47		I have many good qualities.
48		I am satisfied with my achievement so far.
49		I am positive about myself.
50	Self-understanding	I have a clear picture of what I am like as a person.
51		I know my strengths and weaknesses.
52		I have a clear understanding of my own values and principles.
53		I know what I need in my life.
54	Commitment to Self-improvement	I am always motivated to learn.
55		I always keep my knowledge and skills up to date.
56		I look out for new skills or knowledge to acquire.

---



### **About the Authors**

**Ka Hing Lau** is a senior project officer in the Office of Service-Learning at Lingnan University, Hong Kong.

**Robin S. Snell** is an adjunct professor in the Department of Management at Lingnan University, Hong Kong.

Correspondence concerning this article should be addressed to Ka Hing Lau at [khlau2@ln.edu.hk](mailto:khlau2@ln.edu.hk).