THE DEVELOPMENT OF AN INSTRUMENT TO MEASURE PRACTITIONER KNOWLEDGE OF SERVICE-LEARNING

ABSTRACT

The purpose of this study was to develop an instrument that would measure practitioner knowledge of service-learning and to make recommendations based on information gathered in order to guide the statewide service-learning initiatives of Tennessee 4-H Youth Development and the Tennessee Department of Education.

This study utilized a purposefully drawn sample consisting of University of Tennessee Agricultural Extension Service staff who had a 4-H appointment in their job description and participants in the Skills for Action service-learning training offered by the Tennessee Commission on National and Community Service and the Tennessee Department of Education.

The instrument was developed based upon the Points of Light Foundation model of service-learning. The instrument contained three sections. Section one gathered information related to the respondents’ experience with service-learning. Section two measured knowledge of community needs, learning objectives, youth voice and planning,
orientation and training, meaningful action, reflection, evaluation, and celebration and recognition. Section three gathered demographic information.

The following conclusions were based on the findings of this study:

1. Internal consistency of the survey instrument was well above the .66 level set *a priori* in the pilot and sample studies.

2. Internal consistency coefficients dropped dramatically in the test-retest study; with only five of the eight subscales measuring above the .66 level *set a priori*.

3. Stability of the initial survey instrument appears to be low.

4. Tennessee Extension 4-H professionals self-report a slightly higher average number of days of service-learning training than do Tennessee K-12 educators.

5. Construct validity of the survey instrument appears to be low.

6. Tennessee K-12 educators scored statistically significantly higher in all subscales of the instrument than did Tennessee Extension 4-H professionals.

7. Non-community-based advisors scored statistically significantly higher in all subscales of the instrument than did community-based advisors.

8. School-based advisors scored statistically significantly higher in the planning and implementation subscale than did non-school-based advisors.

9. Those who did not volunteer with 4-H Youth Development programs scored significantly higher with regards to the celebration and recognition subscale than did those who volunteer with 4-H Youth Development programs.
Introduction

John Dewey stated that for education “to accomplish its ends both for the individual and for the society it must be based upon experience – which is always the actual life-experience of some individual” (1938, p. 89). To date, many researchers refer back to the work of Dewey, and others, in order to defend and lobby for experiential education. Although the term “service-learning” only emerged in the literature in the 1960’s, the concept of service-learning has been in existence for many years. In addition, Jacoby and Associates wrote, “As a pedagogy, service-learning is education that is grounded in experience as a basis for learning and on the centrality and intentionality of reflection designed to enable learning to occur” (1996, p.9).

As a form of experiential education, service-learning is entrenched in well-established educational and cognitive theories of constructivism, pragmatism, progressivism, and experiential education. These theoretical foundations envelop a wide range of cognitive and affective outcomes for students. The educational domains of experiential learning programs include students’ “intellectual, social, personal, civic, moral, and vocational development” (as cited in Furco & Billig, 2002, p. 27).

Service-learning is a form of experiential learning where students apply knowledge, skills, critical thinking and wise judgment to solve genuine community needs (Toole & Toole, 1994). The practice of service-learning is often closely tied to formal or academic learning environments in school-based and community-based settings. Students in these service-learning programs engage the community in identifying needs, establish learning objectives, empower youth throughout the process, learn about the organization and skills required for serving, conduct the service project, reflect on their
experiences, evaluate the process, and celebrate their successes. Service-learning programs have grown greatly over the past ten years, involving more than six million students at the secondary level alone (Billig & Waterman, 2003).

Through its Communities as Places of Learning Initiative, the Points of Light Foundation sought to bridge the gap between classroom and community learning by working with communities to create service-learning opportunities for youth. Research indicates that actively involving young people in the community and connecting these experiences with the classroom relates positively to young people’s social, personal and intellectual development (Points of Light Foundation, 1997).

The Foundation assisted the National Youth Leadership Council, a national advisory committee of volunteer and nonprofit organizations, and four California sites to develop tools, to assist agencies and schools to create service opportunities for youth. The developments of these resource and training materials helped agencies create service-learning opportunities for youth that enhanced classroom education. The result of this project was the development of an eight-step model for effective service-learning programs (Points of Light Foundation, 1997). The model included community needs, learning objectives, youth voice and planning, orientation and training, action, reflection, evaluation, and celebration and recognition. Understanding this model provides a conceptual framework for the design of the survey instrument.

Given the prevalence of service-learning, it is surprising to see so little research in the field. The vast majority of published service-learning literature includes program evaluations or anecdotal descriptions, and researchers have not established an instrument that measures the knowledge of practitioners with regard to service-learning. (Billig &
Waterman, 2003). Clearly, more rigorous and replicable research is needed in the field of service-learning.

**Purpose**

The purpose of this study was to develop a survey instrument to measure practitioner knowledge of service-learning and to make future recommendations based on information gathered in order to guide the statewide service-learning initiatives of Tennessee 4-H Youth Development and the Tennessee Department of Education. The objectives of this study were to:

1. develop a reliable and valid instrument to measure knowledge of service-learning;
2. examine the relationship between number of hours of service-learning training and score on the measurement instrument;
3. examine the relationship between number of service-learning projects directed and score on the measurement instrument;
4. examine the relationship between selected other demographics and score on the measurement instrument; and
5. make recommendations for future statewide program direction for the Tennessee 4-H program in the area of service-learning.

**Methodology**

**Sample**

The sample utilized in this study included employees of the University of Tennessee Agricultural Extension Service. Specifically, these employees were Extension personnel
with 4-H assignments as a part of their positions. The sample also included a random sample of other educators who have participated in the Lions Quest Skills for Action service-learning training offered by the Tennessee Commission on National and Community Service and the Tennessee Department of Education. This is one of the many trainings that certify high school teachers to teach freestanding service-learning courses in Tennessee.

Through a random numbers generator computer program, 200 were randomly selected for participation in the study. Two hundred participants were selected because that is the approximate number needed for a 95% confidence interval according to Warmbrod (1965). Forty participants were selected to participate in the pilot study after the study sample was selected to eliminate the possibility of members of the pilot group being reselected for participation in the actual study.

A test-retest study was conducted with a purposefully drawn sample of 20 Extension staff members. The purposefully drawn sample was selected based upon the participants’s response rate in the sample study.
Instrumentation

An instrument to measure practitioner knowledge of service-learning was developed by the researcher, based upon the Points of Light Foundation eight-step model for effective service-learning programs. Likert-type scale questions were developed in each of the eight subscales of the model in an attempt to measure practitioner knowledge in each of the eight subscales.

After development of the instrument, a panel of experts examined it in order to establish a high level of content validity. This panel of experts included service-learning professionals and researchers who were internationally known for their reputation in service-learning, including representatives from the Kellogg Foundation, the Points of Light Foundation, private service-learning evaluation and consulting firms, the Tennessee Commission on National and Community Service, and Tennessee 4-H Youth Development staff. All recommended changes from the panel of experts were implemented into the survey.

Data Collection

The survey was conducted by a direct mailing. The survey was mailed to the sample with a cover letter outlining the need for the survey and that the survey was confidential. The point that the survey was voluntary was also stressed. A self-addressed stamped envelope was included for return of the survey. The initial mailing of the pilot test occurred on March 12, 2004, followed by a second letter and survey to non-respondents two weeks later. The response rate for the pilot study was 67.5%. The initial mailing of the sample study occurred on April 19, 2004, followed by a second
letter and survey to non-respondents two weeks later. The response rate for the sample study was 62.5%. The mailing of the test-retest study occurred on May 18, 2004. The response rate for the test-retest study was 65%.

**Data Analysis**

The data were analyzed using the SPSS version 12.0 for Windows. Descriptive statistics including frequencies, percentages, means, modes, standard deviations, range, percents, and frequencies were utilized. Chi Squares, T-Tests, correlation coefficients, and coefficients of stability were used to analyze the relationships in the study.

**Findings**

The first objective sought to develop a reliable and valid instrument to measure knowledge of service-learning. In an attempt to eliminate non-response error, early respondents were compared to late respondents in both the pilot and sample studies. Miller and Smith (1983) indicated that late respondents are often similar to non-respondents. Through T-Tests and Chi Square statistical analysis, late respondents were compared to early respondents, in an attempt to justify generalizing from the respondents to the sample. Statistical analysis failed to produce any statistically significant difference between early and late respondents on all key variables. Therefore non-response was considered to be random in both the pilot and sample study.

The researcher then sought to examine the internal consistency of the instrument. Internal consistency coefficients were calculated using Cronbach’s Alpha. Initial internal
consistency coefficients ranged from 0.76 to 0.87 for the pilot study, from 0.72 to 0.83 for the sample study, both well above the 0.66 level set *a priori*.

Questions were omitted from each subsection in order to develop the most internally consistent instrument. Questions omitted in the pilot study were included in the pilot sample study. Questions omitted in the sample study were included in the test-retest study. Table 36 summarizes the number of questions that were omitted for each of the eight subscales and the final internal consistency coefficients.

Coefficients of stability were then calculated with regards to the test-retest study. Table 37 summarizes the coefficients of stability with regards to each subscale with all

### TABLE 1: Internal Consistency Results

<table>
<thead>
<tr>
<th>Instrument Subscale</th>
<th>Pilot Study</th>
<th>Sample Study</th>
<th>Test-Retest Study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No. of Questions Omitted</td>
<td>Final Internal Consistency Coeff.</td>
<td>No. of Questions Omitted</td>
</tr>
<tr>
<td>Community Needs</td>
<td>1</td>
<td>0.76</td>
<td>3</td>
</tr>
<tr>
<td>Learning Objectives</td>
<td>1</td>
<td>0.84</td>
<td>1</td>
</tr>
<tr>
<td>Youth Voice</td>
<td>2</td>
<td>0.81</td>
<td>5</td>
</tr>
<tr>
<td>Orientation &amp; Training</td>
<td>7</td>
<td>0.89</td>
<td>2</td>
</tr>
<tr>
<td>Planning &amp; Implementation</td>
<td>3</td>
<td>0.89</td>
<td>2</td>
</tr>
<tr>
<td>Reflection</td>
<td>3</td>
<td>0.87</td>
<td>2</td>
</tr>
<tr>
<td>Evaluation</td>
<td>5</td>
<td>0.93</td>
<td>0</td>
</tr>
<tr>
<td>Celebration &amp; Reflection</td>
<td>2</td>
<td>0.92</td>
<td>1</td>
</tr>
</tbody>
</table>
**TABLE 2: Coefficients of Stability on Each Subscale and the Total Instrument Score**

<table>
<thead>
<tr>
<th>Instrument Subscale</th>
<th>Coefficient of Stability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Entire Instrument</td>
</tr>
<tr>
<td>Community Needs</td>
<td>0.68</td>
</tr>
<tr>
<td>Learning Objectives</td>
<td>0.58</td>
</tr>
<tr>
<td>Youth Voice</td>
<td>0.17</td>
</tr>
<tr>
<td>Orientation &amp; Training</td>
<td>0.23</td>
</tr>
<tr>
<td>Planning &amp; Implementation</td>
<td>0.48</td>
</tr>
<tr>
<td>Reflection</td>
<td>0.75</td>
</tr>
<tr>
<td>Evaluation</td>
<td>0.46</td>
</tr>
<tr>
<td>Celebration &amp; Reflection</td>
<td>0.60</td>
</tr>
<tr>
<td>Total Instrument Score</td>
<td>0.38</td>
</tr>
</tbody>
</table>

questions being included, and with regards to each subscale utilizing the most reliable scales from the sample study. While all of the coefficients of stability were positive, only the reflection subscale measured above the 0.70 level set *a priori*. Data regarding the stability of the instrument appeared to be low.

Objective two sought to examine the relationship between number of hours of service-learning training and score on the measurement instrument. Correlation coefficients were calculated between the score on each of the eight subscales on the survey and the number of days of reported service-learning training, as reported in Table 38. According to Davis’ (1971) convention, a negligible positive correlation exists between days of service-learning training and all subscales, excluding the evaluation subscale. A low positive correlation is present between days of service-learning training
and scores on the evaluation subscale of the instrument. Slightly higher positive correlations were observed in the pilot study, but still remained in the low to moderate range.

In order to satisfy objective three, the relationship between the numbers of service-learning projects directed or participated in and the score on the survey was examined, as reported in Table 39. Note that according to Davis’ (1971) convention, a positive negligible correlation existed between number of service-learning projects directed or participated in and score of the planning and implementation, reflection, and evaluation subscales of the instrument. A low positive correlation existed between all other subscales and number of service-learning projects directed or participated in.
TABLE 4: Correlation Between Most Reliable Scales and Number of Service-Learning Projects Directed or Participated In

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Number of Service Learning Projects Pearson Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community Need (n=106)</td>
<td>.110</td>
</tr>
<tr>
<td>Learning Objectives (n=106)</td>
<td>.119</td>
</tr>
<tr>
<td>Youth Voice (n=104)</td>
<td>.127</td>
</tr>
<tr>
<td>Orientation &amp; Training (n=106)</td>
<td>.110</td>
</tr>
<tr>
<td>Planning &amp; Implementation (n=107)</td>
<td>.090</td>
</tr>
<tr>
<td>Reflection (n=107)</td>
<td>.041</td>
</tr>
<tr>
<td>Evaluation (n=105)</td>
<td>.043</td>
</tr>
<tr>
<td>Celebration &amp; Recognition (n=105)</td>
<td>.140</td>
</tr>
</tbody>
</table>

Results in the pilot study ranged from moderate negative correlations to low positive correlations.

Objective four sought to examine differences on subscales among known demographic variables. In the pilot study the only statistically significant differences were observed with regards to occupation and the learning objectives subscale (t = 2.06, df = 24, p = .050), the reflection subscale (t = 2.999, df = 24, p = <.01), and the celebration and recognition subscale (t = 2.302, df = 22.32, p = .031). In all three instances K-12 educators scored statistically significantly higher than Extension professionals in the respective subscales. No other significant differences were observed.
with regards to demographic variables and score in each subscale of the instrument in the pilot study.

Statistically significant differences were observed with regards to occupation and all subscales in the sample study: the community needs subscale \( t = 3.94, df=102.94, p<.01 \), the learning objectives subscale \( t=3.43, df=110, p<.01 \), the youth voice subscale \( t=3.076, df=104.85, p<.01 \), the orientation and training subscale \( t=3.303, df=111, p<.01 \), the planning and implementation subscale \( t=3.029, df=111, p<.01 \), the reflection subscale \( t=4.660, df=108, p<.01 \), the evaluation subscale \( t=3.821, df=109, p<.01 \), and the celebration and recognition subscale \( t=3.337, df=109, p<.01 \). In all cases K-12 educators scored significantly higher than did Extension professionals.

The same trend was observed with regards to community-based advisors and all subscales in the sample study: the community needs subscale \( t = -.4641, df=117.99, p<.01 \), the learning objectives subscale \( t=-3.330, df=118, p>.01 \), the youth voice subscale \( t=-.623, df=114.981, p=.01 \), the orientation and training subscale \( t=-3.692, df=118, p=.000 \), the planning and implementation subscale \( t=-3.002, df=115.681, p<.01 \), the reflection subscale \( t=4.444, df=115, p<.01 \), the evaluation subscale \( t=-3.684, df=116, p<.01 \), and the celebration and recognition subscale \( t=-3.319, df=117, p<.01 \). In all cases non-community-based advisors scored significantly higher than did community-based advisors.

There was only one other statistically significant difference observed in the sample study with regards to demographic variables and scores on individual subscales. Those who did not volunteer with 4-H programs scored statistically significantly higher with regards to the celebration and recognition subscale \( t=-1.977, df=113, p=.050 \).
Correlations among subscale scores and interval scaled demographic variables were also examined. While both negative and positive correlations existed among subscale scores and average volunteer hours, age, and years employed, it should be noted that all were in the magnitude of negligible and low using Davis’ (1971) convention. Similar results were reported in the pilot study, with the exception of age and the community needs subscale, where a negative substantial correlation existed.

Conclusions

The following conclusions were based on the findings of this study:

1. Internal consistency of the survey instrument was well above the .66 level set \textit{a priori} in the pilot and sample studies.

2. Internal consistency coefficients dropped dramatically in the test-retest study; with only five of the eight subscales measuring above the .66 level set \textit{a priori}.

3. Stability of the survey instrument was low.

4. Tennessee Extension 4-H professionals self-report a slightly higher average number of days of service-learning training than do Tennessee K-12 educators.

5. Construct validity of the survey instrument was low.

6. Tennessee K-12 educators scored statistically significantly higher in all subscales of the instrument than did Tennessee Extension 4-H professionals.

7. Non-community-based advisors scored statistically significantly higher in all subscales of the instrument than did community-based advisors.
8. School-based advisors scored statistically significantly higher in the planning and implementation subscale than did non-school-based advisors.

9. Those who did not volunteer with 4-H Youth Development programs scored statistically significantly higher with regards to the celebration and recognition subscale than did those who volunteer with 4-H Youth Development programs.

**Implications & Recommendations**

The most reliable survey instrument should be further developed through research in order to fully develop a valid and reliable instrument. While internal consistency of the instrument was high in the pilot and sample studies, data indicated a low level of validity. Further study is recommended in order to achieve the most reliable and valid instrument. This research may include the further development of questions or rewriting questions that did not score above the .66 level set *a priori*. Future research should engage another panel of experts to fully develop content validity of the survey instrument.

Construct validity appears to be low, however one should also consider the implications if the instrument is measuring accurately. If the instrument is truly measuring accurately then there are implications for the type of training that both K-12 educators and Tennessee 4-H Extension professionals receive. Results in this study indicate that if the instrument is measuring accurately, Tennessee 4-H Youth Development professionals are receiving a less effective type of training than Tennessee K-12 educators. It is recommended that future study be developed to examine the type of trainings offered to both groups, as well as the level at which the trainings are being
offered. Research that examines the type of educational background of each group may directly affect the level at which program officials are able to begin service-learning training.

The most reliable survey instrument should be further studied with known samples in order to fully establish reliability and validity data. Prior to further study, a dynamic service-learning training should be identified utilizing the theoretical framework discussed in this study. The researcher recommends that the survey instrument be administered to the sample prior to this high-quality service-learning training. At the conclusion of the high-quality service-learning training, a post-test should be administered to the same sample. By studying data with groups that are known to receive a high-quality service-learning training, researchers will be able to further develop a valid and reliable instrument.

It is recommended that the most reliable instrument developed in this study be further studied over time with known samples. In this study, respondents were able to self-report the number of training days and the number of service-learning projects directed or participated in. In order to fully establish validity and reliability data, it is recommended that training received and the number of projects directed or participated in be documented over time by researchers. This data should be compared to participant score on the survey instrument. This information should be used to further develop the instrument and add to the validity data of the survey instrument.

It is noted in this study that K-12 educators self-report a smaller average number of days of service-learning training than do Tennessee 4-H Youth Development staff. However, K-12 educators score significantly higher than 4-H Youth Development staff in
all subscales of the instrument. School-based advisors also scored significantly higher in all subscales of the instrument than did non-school-based advisors. The Tennessee state 4-H office should give attention to this study with regards to occupation and whether or not participants are school-based advisors and score on the survey instrument. While care should be given to only generalize the results of this study to the population from which the samples were drawn, it seems clear in this study that of those surveyed, a relationship with a school-based program, whether a K-12 educator and/or a school-based advisor, indicated a higher score in all subscales of the instrument. This may have implications for all community-based organizations. In particular, current service-learning training and resources available to 4-H Youth Development professionals in Tennessee should be evaluated and further developed to provide them with the best possible support from the state program level. Future study should also include researching the effects of work environment and score on the most reliable survey instrument.

Internal consistency data dropped dramatically with regards to the test-retest sample. This decrease would be expected to some degree due to the relatively small, purposefully drawn test-retest sample. Thus, it is recommended that internal consistency be further examined with a larger, truly random test-retest sample.

The instrument was also found to be unstable over time with regards to the test-retest sample. The researcher recommends that the test-retest procedure be conducted with a larger, truly random sample, as opposed to the purposefully drawn sample utilized in this study. The relatively small, purposefully drawn sample that was utilized in this study may have been atypical of the sample utilized in the pilot and sample studies. It is
recommended that further research be conducted with regards to stability of the survey instrument over time, utilizing the most reliable instrument developed in this study and a larger test-retest study.

The researcher also recommends that the Points of Light Foundation eight-step model for effective service-learning programs be subjected to critical scrutiny. While the model appears to have a firm base in theoretical framework, the researcher recommends that further investigation be conducted with regards to the eight-step model. This will further add validity and value to the instrument to measure practitioner knowledge of service-learning.

Through the examination of internal consistency coefficients, the researcher was able to eliminate a total of 16 questions in the sample study. Future research should make a conscious effort to keep the instrument to a limited number of questions, while being careful not to jeopardize the reliability or validity of the instrument. The survey utilized in this study examined 92 variables, with 79 of these being the Likert-type questions measuring the eight subscales. A smaller, more user friendly survey will assist in lowering response error, and could increase response rate.
REFERENCES


National and Community Service Act of 1990.


