

Tools for Assessing Cognitive Outcomes of Experiential Learning

Peggy Fitch & Pamela Steinke

Paper prepared for the 2013 NCA HLC Annual Conference on Quality in Higher Education

The session on which this paper is based introduces the Problem-Solving Analysis Protocol (P-SAP) and the Cognitive Learning Scale (CLS), two instruments that may be used to assess cognitive outcomes of experiential learning related to problem-solving, critical thinking and student perceptions of academic learning. The P-SAP uses a written problem-solving protocol to assess critical analysis and problem-solving. A real-world issue relevant to the course material is presented to students with a set of follow-up questions related to consequences, causes and solutions. The protocols are scored using two global rubrics, one that measures locus of problem-solving and the other that measures complexity of problem-solving. The CLS is a 9-item pretest-posttest that uses a Likert-type rating scale. The pretest CLS uses the stem, “Typically, course requirements that go beyond participation in class and assigned readings:” and the posttest CLS uses the stem, “In this course, course requirements that went beyond participation in class and assigned readings:” The set of statements are the same for the pretest and posttest and include items such as, do/did “help me to spontaneously generate my own examples of principles and concepts I am learning about in class” and do/did “*not* teach me how to apply things I learned in class to real problems” (reverse scored). Both instruments were originally developed for research on the effects of service-learning but have more recently been applied to outcome assessment.

Participants who attend the session will be able to: (1) Identify the strengths, limitations, and appropriate uses of the P-SAP and the CLS; (2) Specify how to score and analyze data generated from the two instruments; (3) Apply the use of the instruments to their home campuses. Both tools are available for use at no charge by contacting the first author at fitchp@central.edu; the authors simply request that those who use them share their findings in order to contribute to the ongoing development of the tools. Participants most likely to benefit from attending the session and reading this paper would be those whose role is to assess relevant cognitive outcomes for undergraduate or graduate students in academic and/or co-curricular programs, particularly those that may involve service-learning or other experiential learning approaches.

This paper extends information presented in the session by including findings from research studies that have used the tools – including available reliability and validity measures – as well as how the tools have been used to assess student learning. Elsewhere the authors have written more extensively about assessment of cognitive

outcomes of service-learning. In the electronic journal *Research and Practice in Assessment* Steinke and Fitch (2007a) review a variety of tools that can be used to assess knowledge application, intellectual development, critical thinking and problem-solving as outcomes of service-learning. Both the P-SAP and CLS are included in this review. Fitch, Steinke, and Hudson (2013) discuss research and theoretical perspectives on cognitive outcomes of service-learning, specifically transfer, problem-solving, and metacognition as cognitive processes related to critical thinking and intellectual development. The P-SAP is included as one of several tools and strategies that can be used to assess service-learning outcomes.

Review of Research Studies Using the Instruments

The authors have used both the CLS and the P-SAP in several previous research studies. The results of the P-SAP reported here involved coding each individual question for both locus and complexity rather than generating a global measure of each. Steinke, Fitch, Johnson, and Waldstein (2002) conducted an interdisciplinary study of service-learning courses across twelve small colleges. Predictors included variables that prior research has identified as best practices in service-learning such as reflection, quality of service-learning placement, community engagement, diversity within the service-learning experience, and student voice/choice. Outcomes included cognitive learning (as measured by an earlier version of the CLS), intellectual development, civic engagement, spiritual and ethical development, and community impact. All five predictors were significantly related to cognitive learning, but much less so for the other outcomes.

Steinke and Harrington (2002) used an earlier version of the CLS as well as an open-ended measure of shared knowledge between the instructor and the student in response to an open-ended protocol. This protocol was a precursor to the P-SAP. The open-ended measure showed good inter-rater reliability between two independent coders with a correlation of $r(183) = .86, p < .001$. Scores on both measures were highly correlated and both were positive indicators of cognitive learning in natural science service-learning courses.

The development of the P-SAP is described in Steinke and Fitch (2003) where results from tests of reliability and validity are reported. Using responses to the P-SAP from a range of courses in different disciplines, this study found good inter-rater reliability between two independent coders with a correlation of $r(23) = .77, p < .001$. An 8-item earlier version of the CLS showed good inter-item scale reliability with $\text{Alpha} = .87$. Further, the P-SAP measures were significantly related to other measures specific to intellectual development and cognitive learning.

Further research on both the CLS and the P-SAP is reported in a paper by Steinke and Fitch (2007b) where they analyzed the results of both instruments as used in three pairs of courses from three different disciplines. Each pair included one service-learning course and one comparable course with a non-service-learning active-learning assignment. Based on the research done previously an item specific to depth of understanding was added to the original 8-item version of the CLS and the final 9-item version of the CLS was used in this study. Both the pretest and the posttest CLS showed good inter-item scale reliability with Alphas = .86 and .83, respectively. On the P-SAP, the percent agreement between two independent coders across the total number of variables coded showed an overall agreement of 84%. The responses of students in service-learning classes were compared to the responses of students in the comparable non-service-learning classes for both measures. Results revealed that students in service-learning classes scored significantly higher than students in comparable classes on the posttest CLS and that the CLS means increased over the semester for service-learning students but decreased for those in comparable classes. The findings were particularly striking because the non-service-learning classes included active-learning projects. Students in service-learning classes also scored significantly higher than students in comparable classes on the P-SAP locus and complexity measures. These results demonstrated consistent findings between direct and indirect measures of cognitive learning although, as is consistent with other findings in service-learning, the effect size for the indirect, self-report CLS measure was larger than the effect size for the direct P-SAP measure. In addition, course grade point was positively correlated with all P-SAP measures and posttest CLS scores were positively correlated with all P-SAP measures except one. These results provided evidence of construct validity for the cognitive skills captured by the P-SAP.

Recent Developments in the Instruments for Use in Assessment

Given that the instruments were both originally developed for research on course-based service-learning, some modifications and further testing have been done and are in progress to adapt the measures to student learning outcome assessment across a range of curricular and co-curricular programs. Institutions that have used the instruments come various states and include representation from community colleges, universities and liberal arts colleges. It is important to the authors that those who use the instruments feel comfortable adapting them to their own assessment needs and program outcomes.

The CLS has remained a 9-item pretest-posttest scale but modifications are being tested on the stem for more flexibility and greater clarity depending on the context in which it is used. For example, some users of the

scale wish to specify the type of course requirements that go beyond participation in the class and assigned reading, or give several examples of what these might include. Those who use the CLS for co-curricular programs have needed to make minor edits to both the stem and the items including substituting the word “program” for “class.” In some cases, it does not make sense to include both a pretest and a posttest so only the posttest is used.

The P-SAP now includes a global scoring rubric for greater ease of use in assessment and is being tested for various ways to integrate it into assessment including within course assignments. The authors have experimented with embedding the instrument into their courses both as a class reflection tool and as a question included in the final exam. This has allowed the instrument to generate pretest and posttest responses by being embedded as a class exercise at the beginning of the course and as a final exam question at the end of the course. Those at other institutions have used it to assess more institution-wide outcomes such as General Education outcomes or outcomes specific to the strategic plan by administering it to incoming first-year students, mid-level students and graduating seniors. In this case students may only need to take it once.

The shared vision of the authors for both instruments is that they continue to develop and that those who use them share what they learn so that others can adapt them to be most effective for assessment. In this way the vision is similar to how the current AAC&U rubrics are being used and adapted on individual campuses. Ultimately, making available a variety of valid, reliable, and flexible assessment tools should contribute to the continued improvement of student learning overall.

References

- Fitch, P, Steinke, P, and Hudson, T. 2013. Research and theoretical perspectives on cognitive outcomes of service learning. In P. H. Clayton, R.G. Bringle, & J.A. Hatcher (Eds.), *Research on service learning: Conceptual frameworks and assessment, Vol. 2A: Students and faculty* (pp. 57-83). Sterling, VA: Stylus.
- Steinke, P., and Fitch, P. 2003. Using written protocols to measure service learning outcomes. In S. H. Billig & J. Eyler (Eds.), *Deconstructing service learning: Research exploring context, participation, and impacts* (pp. 171–194). Greenwich, CT: Information Age.
- Steinke, P., and Fitch, P. 2007a. Assessing service-learning. *Research & Practice in Assessment*, 1(2), 1–8.
Retrieved from <http://www.virginiaassessment.org/rpa/2/Steinke%20Fitch.pdf>
- Steinke, P., and Fitch, P. 2007b. Using goal-based learning to improve cognitive learning outcomes: The case of service-learning. Unpublished manuscript.
- Steinke, P., Fitch, P., Johnson, C., and Waldstein, F. 2002. An interdisciplinary study of service-learning predictors and outcomes among college students. In S. H. Billig & A. Furco (Eds.), *Advances in service-learning research, Vol 2: Service-learning research through a multidisciplinary lens*, (pp. 73-102). Greenwich, CT: Information Age.
- Steinke, P. and Harrington, A. 2002. Implementing service-learning in the natural sciences. [Electronic version] *National Society for Experiential Education Quarterly*, 27 (3), 4-10.