

## **Development of an Artifact-Based Evaluation Framework for Assessing 4-H Learner Outcomes**

### **Abstract**

Effective evaluation requires the selection of appropriate methods to balance rigor and feasibility. Evaluation methods involving surveys and interviews are familiar; lesser known are methods involving the use of participant-generated artifacts. In this article, I share my process for developing an evaluation framework to assess learning outcomes by using artifacts designed and built by young people in 4-H Junk Drawer Robotics programs. Findings demonstrated the potential value of using participant-generated artifacts for outcome evaluation. The process might be replicated in other Extension programs.

**Keywords:** [evaluation methodology](#), [science education](#), [4-H youth development](#)

### **Steven M. Worker**

4-H Youth  
Development Advisor  
University of California  
Division of Agriculture  
and Natural Resources  
Novato, California  
[smworker@ucanr.edu](mailto:smworker@ucanr.edu)

## **Background**

Although a goal of the 4-H youth development program is the improvement of scientific literacy, there is a general lack of systematic outcome evaluation for 4-H science programs (Worker, Schmitt-McQuitty, et al., 2017). Those who conduct evaluation to assess science learning often focus on a narrow range of learning outcomes using survey methodology (Lewis & Worker, 2015). Recently, Worker, Ouellette, and Maille (2017) encouraged Extension to consider a new definition of learning to extend traditional learning indicators and recommended "embedded and authentic evaluation strategies" ("Implications and Recommendations," para. 8).

Effective evaluation requires the selection of appropriate methods that balance rigor and feasibility (Braverman, 2013). Evaluation methods involving surveys, interviews, and focus group interviews are familiar to Extension. Lesser known is the use of participant-generated artifacts—for example, portfolios, presentations, photographs, art pieces, or artifacts built from an engineering design challenge. Artifacts have been underused in qualitative research but are promising because they are nonreactive and grounded in context; that is, they tend not to alter or intrude on, but rather are firmly grounded in, the program (Merriam & Tisdell, 2016). One form of artifact research in educational settings involves repurposing normal day-to-day program activities and using them for evaluation, an approach often described as applying embedded assessment (Wilson & Sloane, 2000) or implementing authentic performance tasks (Wiggins & McTighe, 2005).

## Study Purpose

In this article, I share my process for developing an evaluation framework for assessing learning outcomes by focusing on the final artifacts designed and built by young people in 4-H Junk Drawer Robotics (JDR) programs. The purpose of my study was to determine whether observing artifacts generated by youth participants could serve as a method for assessing learning. The study was approved by the institutional review board of the University of California, Davis. County and participant names used herein are pseudonyms.

Knowing that there is a broad variety of indicators for learning (e.g., knowledge, skills, attitudes, identity), I focused on a particularly salient indicator of learning for science literacy: strengthening of tool competencies. Tool competencies encompass the knowledge, skills, and dispositions needed to use tools, know which tool is appropriate to make a desired modification, and combine several tools together to modify something. Tools (including materials) are cultural artifacts that mediate human cognition and agentic activity and thus fundamentally influence learning and development (Rogoff, 2003; Vygotsky, 1978). The assumption is that resulting artifacts, designed by participants and built using tools, may serve as evidence for learning.

The context for my development of an artifact-based evaluation framework was 4-H science education programs involving use of the 4-H JDR curriculum (Mahacek, Worker, & Mahacek, 2011). JDR advances a design-based science learning pedagogy, an approach to support young people in planning, designing, and making shareable artifacts (Apedoe & Schunn, 2013; Fortus, Dershimer, Krajcik, Marx, & Mamlok-Naaman, 2004; Kolodner et al., 2003; Roth, 2001). One JDR design project is construction of an arm/gripper, an activity from Level 1 Modules 2 and 3, with the challenge being to design and build a freestanding robot arm that uses levers to pick up and move a weight from one spot to another.

## Development of a Pilot Evaluation Framework

I developed an analytical framework for organizing and aggregating indicators of tool competency as manifested in a shareable artifact. The framework was informed by the design challenge from the 4-H JDR curriculum and the engineering design literature. More specifically, the framework addressed two practices of design: (a) generating ideas for how to solve a problem and (b) evaluating potential solutions through prototypes (National Research Council, 2009). With the former, designers innovate to find a solution that will solve a design challenge, whereas with the latter, designers use tools to modify parts and assemble an artifact that will function.

The framework involved four criteria:

- complexity—the total number of parts and the total number of unique parts in the artifact (e.g., five paint sticks would count as five total parts and one unique part);
- innovation—the character (simple, moderate, or complex) of the modification of materials accomplished through the use of tools;
- functionality—the ability or inability of the artifact to function as the participant intended; and
- resolution—verification or lack thereof that the artifact solved the design challenge.

A limitation was that the framework did not assess optimization (i.e., a young person may have used fewer parts more efficiently) and instead privileged experimentation with tools and materials. In other words, assessment of

young people's abilities to engage in engineering design, including the ability to design and build a simple and effective solution, was not included in the framework. I considered this acceptable because I was assessing tool competency and not engineering design abilities per se.

## **Testing of the Pilot Evaluation Framework**

Testing the evaluation framework, and assessing its validity and reliability, required implementing the framework alongside other methods. There were two sites included in my study, Balboa County (seven youths) and Clark County (seven youths). At each site, the program lasted 6–10 hr total over several months.

I employed qualitative multiple-case study methodology (Stake, 2006) using observations of participants, interviews with educators (Seidman, 2013), and focus group interviews with youths (Krueger & Casey, 2015). Observations of participants were documented through written field notes and photographs of 4-H volunteer educators and youths at every session. Data collection occurred in 2014 and 2015. Data analysis began during fieldwork, with analytical memos (Merriam & Tisdell, 2016). Claims were triangulated through multiple sources of evidence.

## **Summary of the Qualitative Case Studies**

I found that educators were using two primary strategies that improved tool competencies. With these strategies, they were allowing for both autonomy and group reflection.

First, educators allocated time for youths to engage in open design and build, where youths took the lead in deciding how to fulfill the design challenge. During these times, the educators supported youths, but did not offer direct instructions. I observed youths using an assortment of tools to modify materials, including hammers, pliers, scissors, glue guns, hand drills, snap punches, and screwdrivers. Guided by the curriculum's design challenge, and mediated by the educator, youth participants had freedom to choose when to use a particular tool and thus were able to make their own decisions regarding selecting a tool, using the tool, and responding to the result. Tool autonomy afforded young people opportunities to grow their tool competencies, by using tools and learning how tools functioned and what their limitations were, all in the service of building a shareable artifact. The 4-H educators did sometimes mediate selection and use of tools. For example, an educator would recommend a particular tool, such as a saw to cut a paint or craft stick or the drill to punch a hole. Other times, an educator disallowed the use of a tool. Additionally, I observed the educators often asking what youths were doing while using tools. The need to justify tool use communicated a group norm that tool use should be task-related and established a purposeful tool use discourse. These approaches helped strengthen the youths' tool competencies.

Second, after youths completed their artifacts, the educators prompted them to share and reflect as a group. Youths communicated challenges they encountered in building their artifacts with regard to use of particular tools. For example, one youth described cutting wood and drilling holes. He conveyed that it was difficult to drill holes exactly where he wanted them but that with practice he was able to accomplish the task (field note, January 29, 2014). This example is representative of youths' sharing how they overcame challenges in using tools and demonstrated growing tool competencies.

## **Balboa Site: Individual Designers**

In Balboa County, each youth designed and built his or her own artifact, and the final artifacts looked different

from one another. For example, one participant constructed an accordion-style arm, and another attempted a grabber claw machine. Over the course of the sessions, all but one of the artifacts became more complex, participants modified more items over time, and participants personalized their artifacts. Examples of data associated with tool competencies are outlined in Table 1.

**Table 1.**  
Examples of Tool Competencies at Balboa Site

Data source	Excerpt
Field notes	The educator [Eugene] provided an introduction to the tools available for youths to use at the second meeting. "Eugene called everyone over to the tool table. He pointed out the tools [and] demonstrated use of the saw and the drills. . . . I observed youths exploring the tools available at the tool table" (field note, 1/29/2014). Additionally, the educator mediated selection and use of tools. "Eugene offered to help Jason with the drill and saw. . . . Eugene asked what [Jason] was going to do next. Jason explained his next task. Jason handed Eugene two paint sticks and Eugene found a drill bit, testing a few different ones" (field note, 3/12/2014).
Postproject interview with educator	Eugene noticed improvement in youths' tool competencies from the first session to the last session. He reflected that his observation of the youths' learning "was really fantastic. . . . you could see it. It was a really nice experience. If we took videos of . . . the last 10 minutes of build time from the very beginning to the very end, it was two different things. When they all started off, they were all walking around, not knowing what to do . . . [They went from that to] at the last meeting, saying 'this doesn't work' and the adult saying 'oh, have you tried this?' and they would go away and try it" (transcript, 4/2/2014).
Postproject focus group interviews with youths	When asked "Did you learn to use a new tool?," Ashley responded, "The pneumatics was new . . . like using the syringes to make [the arm] go up and down." She reported having had the most difficulty with "building the gripper. . . . well, like, I didn't know what to do, so I came back to it later, and then I looked at it, and then I sort of figured it out" (transcript, 3/26/2014).

## Clark Site: Design Teams

In Clark County, youths designed and built their artifacts in design teams, a circumstance that may have afforded more opportunities for members of a group to cross-pollinate ideas. The final three artifacts were similar in appearance. The similarity in appearance may have indicated the flowing of ideas between groups; indeed, I observed participants looking at other groups' devices and informally sharing their thoughts. The artifacts became more complex over time from the first to the last session. Representative excerpts regarding tool competencies are outlined in Table 2.

**Table 2.**  
Examples of Tool Competencies at Clark Site

Data source	Excerpt
Field notes	I observed youths often being more playful with tools at earlier sessions and more task-oriented after a few meetings. For example, "Joyce [the educator] took a minute to share the tools they had available. [She] demonstrated the crop-a-dile, and then they put one at each table. . . . There were youths testing and experimenting with the crop-a-diles" (field note, 10/1/2014). At the last meeting, youths were using

tools with purpose; e.g., Toby using a saw to cut a popsicle stick, Joyce handing Drew tin snips and explaining challenges with cutting tools (field note, 4/1/2015).

Postproject interview with educator Joyce enjoyed "watching the excitement as [the youths] actually got to make their arms move and their grippers grip. . . . Also watching their little brains turn as they're trying to figure out how to make them do what they need to do" (transcript, 5/8/2015).

Postproject focus group interviews with youths All youths communicated that they learned how to build cool things with random materials, a sentiment that represented the intertwined aspects of tool competency, design process, and dispositions. Dex said he learned "definitely how to use random things to build things. . . . it's something—a challenge to do, that's for sure. Like, it's not just something that you can throw together with tape and glue and say it's done. It takes time" (transcript, 4/1/2015).

## Findings for the Evaluation Framework

### Balboa Site

For my final assessment, I used the evaluation framework I developed. Results from the Balboa site are shown in Table 3. Through this evaluation, I determined that the seven artifacts at the Balboa site contained between 14 (least) and 45 (most) total parts and between three (least) and 16 (most) unique parts. Two artifacts exhibited complex material modification, one had moderate modification, and four simple. Four artifacts functioned, one partially functioned, and two did not function. The artifacts created by Isaac and Kayla were the two most divergent (see Figures 1 and 2).

**Table 3.**  
Analysis of Final Arm/Gripper Artifacts (Balboa County)

Criterion	Youth participant						
	Ashley	Jennifer	Ewan	Kayla	Allison	Isaac	Jason
Total number of items	36	27	15	14	20	45	14
Number of unique items	12	7	6	3	7	16	4
Evidence of material modification	Complex; zip ties cut, popsicle stick cut, holes punched in paint stick, popsicle sticks hot glued, hole punched through lid and box	Simple; clothespins separated and glued on base	Simple; holes punched through paint sticks, pieces hot glued	Simple; screws screwed through paint sticks	Simple; paint sticks glued together, holes pocked in sticks, one paint stick cut	Complex; parts not provided by adult used; parts held together by tape, glue, and nails; pieces sawed and cut	Moderate; paint sticks cut and glued together; plastic tubing and syringes glued
Functioned as	Yes	No	No	Yes	Partially	Yes	Yes

intended

Solved design  
challenge

Yes

No

No

No

No

Yes

No

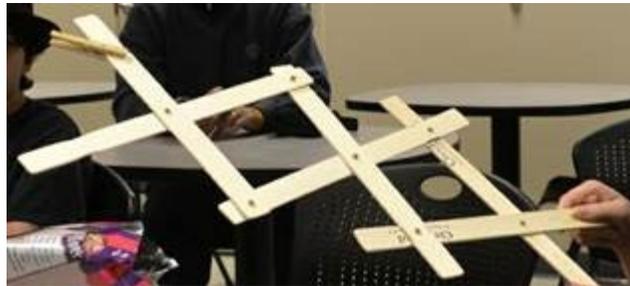
**Figure 1.**

Isaac's Final Artifact



**Figure 2.**

Kayla's Final Artifact



## Clark Site

Results from my use of the evaluation framework at the Clark site are shown in Table 4. The three artifacts there contained between 28 (least) and 41 (most) total parts and between 12 (least) and 15 (most) unique parts. All artifacts provided evidence of complex item modification. All artifacts functioned, although one required a small lift device because the arm could not reach down to table height. As an example, Drew and Glen's final artifact is shown in Figure 3.

**Table 4.**

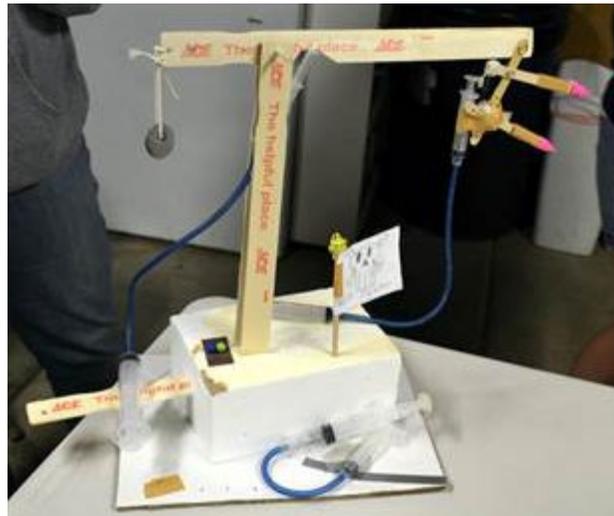
Analysis of Final Arm/Gripper Artifacts (Clark County)

Criterion	Youth participant team		
	Drew and Glen	Cooper and Dex	Jack, Greg, and Toby
Total number of items	28	34	41

Number of unique items	13	12	15
Evidence of material modification	Complex; holes punched in paint sticks, holes sawn in paint sticks, paint sticks sawn, pieces hot glued, pieces taped or held together with string, objects punched into Styrofoam	Complex; holes punched in paint sticks, holes sawn in paint sticks, paint and popsicle sticks sawn, pieces taped or held together with string, objects punched into Styrofoam	Complex; holes punched in paint sticks, paint sticks sawn, pieces taped and glued together, Styrofoam punched through
Functioned as intended	Yes	Yes	Yes
Solved design challenge	Partially, did not fully reach table	Yes	Yes

**Figure 3.**

Drew and Glen's Final Artifact



## Conclusions

Comparing both methods—the qualitative case studies and evaluation framework—revealed how each method illuminated various perspectives on tool competencies. The case studies provided rich descriptions of tool use over time from the perspectives of the researcher (through observations), the educators (through interviews), and the youths (through focus group interviews). Together, these qualitative methods offered insights into intertwined outcomes of tool competencies, design practices, and persistence. However, I expended extensive effort to collect, analyze, and interpret the qualitative data. The evaluation framework provided a quantifiable summary of tool competencies. The evaluation framework, though, was a point-in-time snapshot and did not reveal educators' strategies for promoting tool competencies or young people's experiences using tools. Both methods served a role in evaluating program outcomes. Evaluation is stronger through methodological pluralism. Nonetheless, the evaluation framework may be a strong option when time or resources for evaluation are limited.

My study reinforces the value of extended indicators of learning and the potential for using participant-generated artifacts for outcome evaluation. I was able to see cognition manifested in the 10 shareable artifacts, and with the aid of a research-based framework, organize data from each artifact into a useful and comparable format. The resulting data may be shared as evidence of outcomes in a 4-H science program or applied formatively to future iterations of the program to generate improvements. Future evaluation of JDR Level 1 Modules 2 and 3 might be conducted with the framework and without the need for observations of participants, individual

interviews, or focus group interviews.

## Recommendations

The process I used—(a) identifying participant-generated artifacts to use as evidence for learning, (b) developing an assessment framework with criteria based in the literature, (c) pilot testing the framework alongside other evaluation methods to establish the framework's validity, and (d) sharing the framework as an evaluation instrument—might be replicated in other settings. In this way, program outcomes may be assessed without placing extra burden on the participants, and the assessment may be more sensitive to program activities. Evaluation relying on participant-generated artifacts helps promote an expanded conception of learning as participation in communities while moving Extension professionals beyond a "pre-post survey" approach. One should realize, however, that using participant-generated artifacts requires a substantial initial investment of time to realize future benefit. Additionally, as with all evaluation methodology, evaluation of participant-generated artifacts has limitations that must be considered, including potential issues of reliability and validity.

## Acknowledgments

I extend my appreciation to Cynthia Carter Ching, Lee Martin, and Tobin White for their guidance on the dissertation research from which this article was developed. I thank the editor, anonymous reviewers, and Alexa Maille, whose comments helped improve and clarify this manuscript.

## References

- Apedoe, X. S., & Schunn, C. D. (2013). Strategies for success: Uncovering what makes students successful in design and learning. *Instructional Science, 41*(4), 773–791.
- Braverman, M. T. (2013). Negotiating measurement: Methodological and interpersonal considerations in the choice and interpretation of instruments. *American Journal of Evaluation, 34*(1), 99–114.
- Fortus, D., Dershimer, R. C., Krajcik, J., Marx, R. W., & Mamlok-Naaman, R. (2004). Design-based science and student learning. *Journal of Research in Science Teaching, 41*(10), 1081–1110.
- Kolodner, J. L., Camp, P. J., Crismond, D., Fase, B., Gray, J., Holbrook, J., . . . Ryan, M. (2003). Problem-based learning meets case-based reasoning in the middle-school science classroom: Putting learning by design into practice. *Journal of the Learning Sciences, 12*(4), 495–547.
- Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research* (5th ed.). Thousand Oaks, CA: Sage.
- Lewis, K. M., & Worker, S. M. (2015). Examination of attitude and interest measures for 4-H science evaluation. *Journal of Extension, 53*(3), Article 3RIB4. Available at: <https://www.joe.org/joe/2015june/rb4.php>
- Mahacek, R., Worker, S., & Mahacek, A. (2011). *4-H Junk drawer robotics curriculum*. Chevy Chase, MD: National 4-H Council.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation* (4th ed.). San Francisco, CA: Jossey-Bass.
- National Research Council. (2009). *Engineering in K-12 education: Understanding the status and improving the*

*prospects*. Washington, DC: The National Academies Press.

Rogoff, B. (2003). *The cultural nature of human development*. Oxford, England: Oxford University Press.

Roth, W-M. (2001). Learning science through technological design. *Journal of Research in Science Teaching*, 38(7), 768–790.

Seidman, I. (2013). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th ed.). New York, NY: Teachers College Press.

Stake, R. E. (2006). *Multiple case study analysis*. New York, NY: The Guilford Press.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Wiggins, G. P., & McTighe, J. (2005) *Understanding by design* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Wilson, M., & Sloane, K. (2000). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, 13(2), 181–208.

Worker, S. M., Ouellette, K. L., & Maille, A. (2017). Redefining the concept of learning in Cooperative Extension. *Journal of Extension*, 55(3), Article 3FEA3. Available at: <https://www.joe.org/joe/2017june/a3.php>

Worker, S. M., Schmitt-McQuitty, L., Ambrose, A., Brian, K., Schoenfelder, E., & Smith, M. H. (2017). Multiple-methods needs assessment of California 4-H science education programming. *Journal of Extension*, 55(2), Article 2RIB4. Available at: <https://www.joe.org/joe/2017april/rb4.php>

---

Copyright © by *Extension Journal, Inc.* ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the Journal Editorial Office, [joe-ed@joe.org](mailto:joe-ed@joe.org).

If you have difficulties viewing or printing this page, please contact [JOE Technical Support](#)