SHHH…DON'T USE THE 'A' WORD: EMBEDDED ASSESSMENTS OF YOUTH LEARNING IN INFORMAL LEARNING ENVIRONMENTS

By Melissa Koch, William R. Penuel, Torie Gorges, and Geneva Haertel
SRI International

Finding appropriate ways of assessing learning outcomes of participants in informal learning environments is both a challenge and an important opportunity. The Build IT project addressed the challenge of assessing youth learning in a way that was consistent with the culture and goals of an informal learning environment. Build IT, a problem-based curriculum with performance tasks for IT fluency assessment, encourages middle school girls to develop IT fluency, interest in mathematics, and knowledge of IT careers. This study's focus on the process of design and enactment provides a window on how informal learning programs like Build IT can address the challenges of developing and implementing valid assessments of student learning that fit well in the informal learning environment and have sustainable use by practitioners.

A Focus on Assessment in Informal Science Learning Environments

Informal learning environments play an important role in providing opportunities for youth to learn STEM (science, technology, engineering, and mathematics) concepts and skills (Falk, 2001). Consequently, many practitioners and researchers believe they must find appropriate ways of assessing learning outcomes of participants in informal programs (Allen, 2002; COSMOS, 1998; Falk & Dierking, 2000; Martin, 2004). Yet, most informal science programs still face serious challenges in documenting and presenting evidence of participants' learning (Allen, 2002; Falk & Dierking, 2000). To complicate matters, the word "assessment" often has negative associations with formal learning environments among youth and program leaders. If used as tests are in schools, assessments can undercut youth interests.

The focus of this paper is on how one informal science learning project, Build IT, addressed the challenge of assessing youth learning in a way that was consistent with the culture and goals of an informal learning environment. Build IT is an after school and summer youth-based curriculum for middle school girls to develop IT fluency, interest in mathematics, and knowledge of IT careers. Developed by SRI International (SRI) and Girls Incorporated of Alameda County (Girls Inc.), Build IT is a problem-based curriculum that capitalizes on girls' interest in design and communication technologies and incorporates performance tasks for IT fluency assessment. For Build IT, the key assessment challenge was developing sources of evidence about girls' IT interests and fluencies (National Research Council, 1999a) in ways that were consistent with Girls, Inc.'s hands-on, youth development approach to programming.

In the development of Build IT, SRI and Girls Inc. collaborated to create assessments that assess learning and formative evaluation tools that document learning opportunities.

Embedded in the curriculum are three types of "performance tasks" which allow girls to demonstrate what their learning: activities in which girls apply the enduring understanding they've learned to a new context; guides for group discussion and reflection on concepts; and Family Tech Nights where girls present what they've learned to their families and communities. While these embedded assessments are the focus of this paper, the suite of assessment and formative evaluation tools used by staff includes surveys to measure girls' interests, skills, and conceptual understanding of fundamental IT concepts and self-evaluation tools for staff to gauge youth learning opportunities and program effectiveness. The assessments were designed, enacted, and refined to fit within an informal learning environment and have sustainable use by practitioners.

Theoretical Frameworks

Two theoretical frameworks guided our assessment design. *Evidence-centered assessment design* informed the overall design of assessments, and *expectancy-value theory* within developmental psychology informed ideas about the kinds of assessment tasks that would encourage and sustain youth interest in IT.

The evidence-centered assessment design approach emphasizes the need for learning assessments that elicit evidence that support valid inferences about student learning (Messick, 1992; Mislevy, Steinberg, & Almond, 1999). A core idea is that valid inferences about progress toward learning goals depend on the development of a coherent, well-supported assessment argument about what students know and can do. That argument must be backed by evidence that includes systematic analysis of student performance on tasks designed to elicit targeted student knowledge and skills (Mislevy, 1994). In evidence-centered designs, improving educational assessment turns on the ability of developers to articulate a model of what students should know and be able to do and to identify the tasks that will reveal whether or not students have mastered a target set of knowledge and skills (Mislevy, Steinberg, Almond, Haertel, & Penuel, 2003).

Studies of classroom-based assessment suggest that to be effective, assessments must support student motivation to learn (Black & Wiliam, 1998); and so we used girls' expectations of their own success in IT and the value they place on IT as a second guide to assessment design (Eccles, 2005; Zarrett, Malanchuk, Davis-Kean, & Eccles, 2006). According to the Eccles expectancy-value model, a girl's performance, task choice and persistence are directly linked to two primary components: a girl's expectations for success and the perceived value of the options, which includes girl's interest. Expectancy and value have consistently been shown to predict both academic success and occupational choices (Eccles, 1986; 1987; 1994; 2007; Zarrett et al., 2006). The format and content of the embedded assessments in Build IT are guided by Eccles expectancy-value model: girls' interests in and values of the tasks and careers are encouraged while building girls' successful experiences with mathematics and IT.

Methods for Curriculum and Assessment Development

The evidence-centered design approach and expectancy-value model guided the team in conceptualizing and designing the assessments. To guide the iterative development of the curriculum, the team used two development techniques: *Understanding by Design* (UbD)

and user-centered design. Wiggins and McTighe's (1998) UbD approach provided a structure for the curriculum and assessment developers to identify the enduring understandings, determine what is evidence of learning, and create the activities, including the embedded performance tasks used for assessment. As an approach to curriculum development, its steps mirror those of the evidence-centered design approach, in that developers must first begin with a model of student learning and then proceed to develop tasks that reveal students' level of mastery with respect to learning goals.

The developers also used user-centered design (Cooper, 1999; Nielsen, 1994) techniques to co-develop and iterate with the program leaders and participating girls at Girls Inc. as they implemented the curriculum. The development team's ongoing interactions with users included early feedback from program leaders on drafts of the curriculum and assessments prior to implementation, regular reports from girls and program leaders on their experiences during implementation, and observations and interviews with staff and girls during implementation. This iterative, user experience-based process enabled the development team to refine the materials so that the IT concepts and skills were understood by the program leaders, were engaging for the girls, and promoted girls' and program leaders reflection on girls' learning through the performance tasks, discussions, and Family Tech Nights. The user-centered design techniques helped realize the goal of developing assessments that foster and sustain, rather than undercut, girls' interest in participating in the program.

A Formative Analysis of Build IT's Assessments

To understand the use of assessments in the Build IT program, the team addressed the following two questions:

1. How did the project's use of ***UbD and user-centered techniques*** support and/or hinder the development of assessments that foster girls' interest in IT?
2. How did the program leaders ***enact the assessments***, and how did they transform those assessments in the process of enactment?

The formative evaluation, using a mixed methods approach, was the primary source of data for the study. The goal of the formative evaluation was to document the development, enactment, and revision of the curriculum and performance assessments to foster girls' learning and program improvement. Data were collected using a combination of open-ended observations, semi-structured interviews, and reviews of artifacts created by girls and program leaders. Information gathered from interactions between the curriculum developers and program leaders, as well as suggestions made by the program leaders for curriculum improvements during professional development sessions and curriculum meetings, were sources of data. Data were analyzed in two ways: informally as part of regular meetings with SRI development staff on the project, evaluators communicated their findings to curriculum developers, so that they could respond in a timely manner to challenges the program leaders were facing, and formally, at the end of each year of programming. For formal data analysis, evaluators used both an emergent coding scheme in the tradition of grounded theory (Charmaz, 1983) and coding for pre-defined themes, based on program goals. Additionally, formative evaluators collected a running list of suggestions made by staff and girls throughout the year. Two evaluators

analyzed all data and came to agreement on the conclusions drawn. Occasionally, when we found that we did not have sufficient data to draw conclusions, we conducted additional, informal interviews with Girls Inc. staff.

Results

To develop the curriculum and performance assessments, the team married ***UbD and user-centered techniques***. The team began, as directed by UbD, by defining the enduring understandings and performance tasks goals. The team then divided into two groups: curriculum development and assessment development. These teams reconvened to review the curriculum activities and performance assessments each had created and to bring them into alignment. The division of teams was a deviation from the UbD model of development, but tight project deadlines necessitated that work be conducted in parallel. At the same time, the added step of aligning the activities forced the team members to make sure that the curriculum provided opportunities to develop enduring understandings to be assessed and that the assessments did not reach beyond what the curricular activities could support.

Once a draft of the curriculum and assessments were ready for review, they went through two cycles of edits and revisions with staff at Girls Inc. prior to implementation: one review prior to the professional development and another revision based on feedback during the professional development. While UbD doesn't provide guidance on how to iterate and redesign curriculum and assessments once they are taught, user-centered design provides a structure for ongoing iteration with users, in this case the Girls Inc. staff and participants. Interviews and observations conducted by the formative evaluation team and conversations among the Girls Inc. staff and the development team pointed out strengths and weaknesses in the design of the performance assessments. A key struggle the team faced in incorporating user input was in reconciling youth development perspectives with opportunities for students to learn and reflect on enduring understandings for technology fluency.

In the course of ***enacting the assessments***, program leaders transformed the assessments and pushed for revisions that addressed specific challenges related to girls' interests. For example, at the beginning, we called the tasks "performance assessments," but the leaders regarded the term with suspicion because of its school connotation. We changed the name to "performance tasks" to accommodate them; program leaders felt that the new name aligned well with youth development goals and the curriculum activities. Specific formats for performance tasks that worked well also emerged as part of enactment: tasks where girls gave each other scaffolded feedback in small groups and could then act on this feedback to improve their work worked well for the leaders and the girls. Performance tasks that were directly connected to activities, such as a culminating projects, rather than another activity where the girls applied what they learned, were also successful. In all types of performance tasks, reflection formats are heavily used in the structure of the performance tasks, consistent with research that shows that providing students with time to reflect on their learning and revise their ideas improves learning (National Research Council, 1999b).

One of the unresolved challenges pertains to how leaders see the assessments and understand their use. Now in the third year of the project, each of the six units has two to three embedded performance tasks that the leaders use. However, what the leaders do with the information and whether or not they treat the performance tasks as assessment opportunities varies by leader and situation. Some simply view the performance tasks as activities in the curriculum, while others recognize their special role in providing evidence to support inferences about what girls have learned. In some sense, the fact that assessments for some are "fully embedded" represents a success, but from the standpoint of evidence-centered assessment design, we still need to develop ways to help program leaders use their observations of what girls learn to develop coherent, well-supported arguments about girls' learning.

Educational or Scientific Importance of the Study for Discussion

Programs, like Build IT, that have developed assessments and a process for assessment refinement in collaboration with users can shed light on how to develop, implement, and sustain the use of these assessment tasks in informal learning environments. Through the development process, the team learned how to create assessments that work well in youth development settings. Program leaders reactions to and implementations of these assessments guided the development team in making refinements to the assessments. When implemented well, these embedded assessments, as well as the activities leading up to them, encourage program leaders to move beyond "making cool stuff" to foster girls' learning new skills and gaining understanding of IT concepts. The ability of staff to re-use these assessments relies on the ability of the curriculum and professional development to clearly communicate the skills and conceptual understandings in a youth development format that the program leaders can adopt and sustain. This study's focus on the process of design and enactment provides a window on how informal learning programs can address the challenges of developing and implementing valid assessments of student science learning that fit well in the informal learning environment and have sustainable use by practitioners.

References

Allen, S. (2002). Looking for learning in visitor talk: A methodological exploration. In G. Leinhardt, K. Crowley & K. Knutson (Eds.), *Learning conversations in museums* (pp. 259-304). London: Lawrence Erlbaum Associates.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.

Cooper, A. (1999). *The Inmates Are Running the Asylum: Why High-Tech Products Drive Us Crazy and How to Restore the Sanity*. Indianapolis, IN: Sams Publishing.

COSMOS. (1998). *A report on the evaluation of the National Science Foundation's Informal Science Education Program*. Washington, DC: National Science Foundation.

Eccles, J. S. (1986). Gender roles and women's achievement. *Educational Researcher., 15*(6), 15-19.

Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly, 11*, 135-172.

Eccles, J. S. (1994). Understanding Women's Educational and Occupational Choices: Applying the Eccles et al model of achievement-related choices. *Psychology of Women Quarterly, 18*(4), 585-609.

Eccles, J. S. (2005). Studying Gender and Ethnic Differences in Participatin in Math, Physical Scienc, and Information Technology. In J. E. Jacobs & S. D. Simpkins (Eds.), *Leakds in the Pipeline to Math, Science and Technology Careers* (Vol. Number 110, pp. 55-88). San Francisco, CA: Jossey Bass.

Eccles, J. S. (2007). Where are all the women? Gender differences in participation in physical science and engineering. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science?* (pp. 199-210). Washington, DC: American Psychological Association.

Falk, J. H. (2001). *Free-Choice Science Education: How we learn science outside of school*. New York City: Teachers College Press.

Falk, J. H., & Dierking, L. D. (2000). *Learning from museums: Visitors experiences and the making of meaning*. Walnut Creek, CA: AltaMira Press.

Martin, L. (2004). An emerging research framework for studying informal learning schools. *Science Education, 88*(supplement 1), S71-S82.

Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika, 59*, 439-483.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design: A brief overview*. Princeton, NJ: Educational Testing Service.

Mislevy, R. J., Steinberg, L. S., Almond, R. G., Haertel, G. D., & Penuel, W. R. (2003). Improving educational assessment. In B. Means & G. D. Haertel (Eds.), *Evaluating educational technology: Effective research designs for improving learning.* (pp. 149-180). New York: Teachers College Press.

National Research Council. (1999a). *Being fluent with information technology*. Washington, DC: National Academies Press.

National Research Council. (1999b). *How people learn: Brain, mind, experience*. Washington, DC: National Academy Press.

Nielsen, J. (1994). *Usability Engineering*. San Francisco: Morgan Kaufmann.

Wiggins, G., & McTighe, J. (1998). *Understanding by Design*. Alexandria, VA: ASCD.

Zarrett, N., Malanchuk, O., Davis-Kean, P., & Eccles, J. S. (2006). Examining the gender gap in IT by race: Young adults' decisions to pursue an IT career. In J. M. Cohoon & W. Aspuray (Eds.), *Women and information technology: Research on underrepresentation* (pp. 55-88). Cambridge, MA: The MIT Press.