Design Principles for Assessing Learning in Learning Ecosystems: Fostering Productive Interactions around Digital Badges

Rebecca C. Itow and Daniel T. Hickey

Abstract: The Design Principles Documentation (DPD) Project has followed the development of thirty digital badging projects as they moved from their intended practices outlined in their initial proposals to their enacted practices constrained by their audience and the badging platform. The project then documented which practiced were formalized when the badge systems were brought to users. The DPD Project's immediate goal was to capture the knowledge that emerged in this progression. The result was the derivation of ten design principles for designing assessments in digital badge systems that are relevant beyond badging systems. The broader goal of this work was to study the way in which assessment practices impact the *learning ecosystem.* A scholarly review of the literature related to the design principles revealed aspects of assessment functions that badging projects – or anyone designing assessments for a learning system – should consider as they make decisions about the kind of learning they wish to foster.

Digital badges are web-enabled tokens of accomplishment that contain specific claims and detailed evidence about learning. Because they contain this information and can be readily accumulated and shared, they can work quite differently than traditional grades, certificates, and transcripts (Casilli & Knight, 2013). When the MacArthur/Gates Digital Media and Learning (DML) initiative launched the Digital Badges for Lifelong Learning competition in 2012, a community around digital badges began to form, and efforts spanning a range of educational settings began to emerge. All of these efforts to use digital badges fundamentally recognize some kind of learning, be it informal, vocational, or formal scholastic work. But if one is going to recognize learning, there almost always is some kind of assessment to validate that recognition. Learning system developers must be acutely aware of how their choices impact learning if they are going to make claims about what learners can and cannot do as a result of engaging with their content and activities.

While each of the thirty DML awardees were charged with using digital badges, their task was really to build a learning ecosystem, and designing the activities and assessments that formed productive and dynamic educational spaces became the projects' biggest challenge. It

became quickly apparent that designers needed to shift their focus beyond the digital badges to the interactions that occurred around and on the pathway to earning each badge. Both designers and beta users wanted the badges to *do something*, and many of the projects realized that meant designing activities that engage learners in productive discourse that is both disciplinary and has applicability beyond the badge system. In order to do this, each project made choices about what they would assess (specific hard and soft skills) and how they would assess it. Each project's choices regarding their assessment practices directly impacted learners' engagement and discourse within the environment, which in turn impacted the tone and form of the learning ecosystem.

While assessments can serve both summative and formative functions, the summative functions have the potential to undermine the formative functions. Many badging projects found that as they moved forward and their learner populations grew, they needed to find a balance between summative and formative functions of assessment so that many learners are given appropriate feedback in a reasonable amount of time. Many projects also found that their users valued broadly applicable "soft skills" over the domain specific "hard skills;" this posed an interesting challenge in choosing assessment practices, because they found that soft skills like collaboration and leadership are difficult to define and prove problematic when designing assessments.

Assessing the range of skills valued by both the projects and the learners became a task of balancing the amount and quality of formative feedback a learner received with the necessary summative assessments needed to make claims about learning outcomes. As a result of finding this balance, transformative functions of assessment (Torrance, 2012) emerged in some of the projects. These assessment practices transformed the learning ecosystem into dynamic spaces where learning and assessment were inseparable, and sophisticated discourse and reflection occurred around complex concepts. This paper will outline the general assessment design principles and present three comparative examples of how some of those principles were used to develop transformative assessment practices.

Developing Design Principles for Assessing Learning in Digital Badge Systems

When the *Badges for Lifelong Learning* competition was launched, the Design Principles Documentation (DPD) Project team was charged with the task of following the thirty projects awarded funds to build digital badge systems that encourage learning in a variety of formal and

informal settings. In order to carefully analyze the different aspects of the badging systems and their development over time, we developed the DPD Project and focused our questions and research around emerging practices in four categories of badging practices: recognizing, assessing, motivating, and researching learning with digital badges. Carla Casilli observed that "[r]egardless of where you start, it's more than likely you'll end up somewhere other than your intended destination" (Casilli, 2012)[1] when designing a badge system. With this in mind, the DPD team concerned itself with the emerging informal knowledge that developed as projects moved from intended practices outlined in their initial proposals to enacted practices that developed as the projects matured and systems were implemented with learners. This paper focuses on the assessment strand of this research.

Existing Research on Digital Badges

Efforts throughout the badging community have been made to track and document different aspects of badge system development. Particularly important to this work was Sheryl Grant and Kristan Schwago's *Annotated Research Bibliography* (2013).[2] Their collection of badges-related literature coincided nicely with our search for literature related to badging practices. In October 2013, Sheryl Grant also identified five "buckets" that describe the different places from which badge systems start, and explains how they may move forward with the resources they possess. [3]

Carla Casilli has created a series of concept maps that help digital badge designers understand the components that make up a badge system. One particularly useful map outlines the components needed to develop trust in a system[4]. This map focuses on the notion of "credibility" (May, 2012)[5]. While face validity is often seen as weak evidence, in a digital badging system, the type and amount of credibility outsiders perceive about a system holds great weight and influence in its successful implementation. Credibility is established when the seven related "trust components" Casilli suggests are present.

In Summer 2013, Mozilla teamed up with the MacArthur Foundation and the City of Chicago to create the *Chicago Summer of Learning*, which provided opportunities to Chicago youth to participate in interest-driven learning pathways. A great deal of effort was put into

[1] http://tinyurl.com/Casillistart
[2] https://www.hastac.org/digital-badges-bibliography
[3] http://tinyurl.com/grantbuckets
[4] http://tinyurl.com/Casillitrust
[5] http://tinyurl.com/casillivalidity1

designing environments that fostered meaningful interactions and connected learning (Ito et al., 2013) around the badges.

The DPD Project aims to extend and build upon the work currently being done by others. This research differentiates itself from other work by focusing in on these four strands of learning. This research (a) traced intended, enacted, and formal assessment practices across the thirty projects, (b) derived ten general principles for assessing learning with digital badges, (c) connected these principles to relevant aspects of project contexts, and (d) connected these principles to relevant external research and resources to help projects be more systematic about their assessment choices. The DPD project also ended up connecting projects with similar practices, which resulted in powerful and productive discourse across the projects.

Our derivation of badge design principles and reflexive identification of outside literature focus directly on the ideas and insights most relevant to the badges initiative. Rather than summarizing the vast literature on assessment, the literature was reviewed and the design principles were derived in the same way evidence in a trial must be directly relevant to the case and the question at hand (Maxwell, 2006). Both the literature review and the design principles relate to the enacted practices that emerged as badge systems were implemented. Because there is almost no literature on assessment for badges and a vast literature on assessment, a reflexive and recursive review of the relevant literature and other resources was needed to inform these efforts.

This "conceptual" (rather than "foundational") approach is called for by the groundbreaking nature of this initiative. Teasing apart the areas of recognizing, assessing, motivating, and studying learning and addressing the tensions within and between these four areas serves to highlight the incredibly complex problems in educational reform and begin rectifying some of them. The review and design principles pragmatically review prior research and explore the value of newer paradigms based on new newer "social" theories of learning. These new theories are infused throughout the DML initiative and are embodied in the writings DML scholars and leaders like John Seely Brown (e.g. Brown et al. 1989; 2002; 2008;), Mimi Ito (Ito, et al. 2005; 2009), and Connie Yowell (Yowell & Smylie, 1999).

These social theories are being widely adopted, but many educational innovators embrace social theories of learning while using traditional practices to assess learning. This research aims

to highlight the impact assessment choices have on learning so that educational innovators have a better understanding of the kinds of claims that can be made with those choices.

Deriving principles through design-based research. While carried out as design ethnography, this research was inspired by current design-based research (DBR) methods. DBR builds "local theories" through systematic iterative design in the context of implementations (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003) with particular attention to relevant aspects of particular educational contexts. To help capture the emergence of local theories and relevant aspects of context, this research reflexively documented more general and more specific ways of assessing learning in the context of digital badges. *Meta-principles* are very general theoretical statements that reflect how different "grand theories" of knowing and learning (i.e., behavioral, cognitive, or sociocultural) lead to different conceptualizations of the role of digital badges. *General principles* are the guiding principles behind badge system developers' designs. These general principles are broad, and address the projects' general practices for recognizing, assessing, motivating, and studying learning. *Specific practices* – labeled *appropriate practices*, explained below – reflect how the general design principles were enacted in light of specific constraints of the individual content development efforts. These practices, along with the *specific features* that emerged in that context are invaluable for helping similar projects appreciate how the general principles might be enacted and further refined in their own contexts.

Extracting "appropriate practices" for designing badge systems. Analyses of proposals and subsequent interviews with project stakeholders identified over 100 enacted assessment practices across the thirty projects. The thirty interviews revealed that projects needed to make substantial changes to their initial designs when they began working within the badging platforms in their particular settings. Doing so uncovered aspects of platforms and settings that impact how (and if) initial designs were enacted. These practices were organized into ten assessment design principles, which remained usefully linked back to the specific practices and features. This process was followed by a half-day workshop where stakeholders could review our characterizations of the projects' assessment practices, refinement of the general assessment principles, as well as a recategorization of projects' practices. These practices are documented in the DPD project database[6], to which all of the projects have access. These practices have been labeled *appropriate practices* rather than best practices, as the

[6] http://iudpd.indiana.edu/

appropriateness of a specific practice depends on the context in which it is used. The appropriate practices contained in the ten principles for badging practice are open to the public so that new badging systems can use them to inform their decisions around assessment. This information is now being disseminated back to projects and to the broader badging community via blogs and at www.workingexamples.org where it will continue to evolve.

Resulting design principles. The general design principles for assessing learning in digital badge systems are a result of this process. They are ordered by prevalence among the thirty projects. The first principles are employed by almost all of the projects, while the last principle is used by only three. Badge system developers must consider what kind of learning they want to recognize, and how the assessment principles they use will impact the learning ecosystem.

The development of these principles has led to the need to attach them to the scholarly assessment literature. By bringing in the relevant research literature, both existing and new projects can make informed decisions about their assessment practices, and the DPD project can make recommendations of aspects to consider when employing particular principles. What follows are (a) the ten general design principles and the number of projects who use each principle, (b) the more specific principles under each design principle, (c) one example of relevant research associated with the principle, and (d) an example of a specific practice (also see Table 1 at the end of this document). This is then followed by three comparative examples of the principles in practice.

Use leveled badge systems. Nearly all (twenty-nine) of the thirty projects included some kind of "leveling" system that learners would move through as they practiced new skills, as opposed to a "flat" system where all badges have equal value. Sixteen projects created what we deemed "competency levels," ten used "meta-badges," and three formed hierarchical categories of badges. Projects using benchmark assessments to promote mastery of a specific skill would do well to learn from Bulkley et al.'s (2010) research on Philadelphia schools, finding that while the benchmark assessments revealed general categories that needed to be retaught, the assessments were not designed in such a way that a teacher could use the incorrect answers to understand mistakes in student thinking.

BuzzMath[7] is a competency-based badges system that focuses on mathematics for middle school grades. Learners complete short sets of problems within a particular competency, and earn small badges for passing each of those sets. These smaller badges are not necessarily ones that learners would push out to their backpacks or share with friends, but they serve as markers of progress in mastering the targeted competency. Once a specified number of these smaller badges have been earned, learners can attempt to earn the larger badge that represents mastery of the targeted concept. In the mastery-learning orientation of the *BuzzMath* project, badge levelling allows both learners and educators to identify mistakes and go through a reteaching/learning process before a student earned the next level of a badge.

Align assessment activities to standards. Twenty-six of the projects aligned their projects to existing standards. These standards varied from national and state standards to internal standards set by the parent organizations of the projects. Ten projects used internal standards, seven used national/state standards, and nine used the Common Core State Standards. Darling-Hammond (1997) discussed the need to raise standards and the system in which they are employed to support teaching and learning.

Makewav.es' project, *Supporter 2 Reporter (S2R) Medals*[8] teaches youth to write through sports journalism. They aim to focus sports-minded youth's energies toward writing and participation in their sport of interest. As will be elaborated below, the learners create portfolios to demonstrate growth and to hold final artifacts. There is already a large community of teachers within the Make Waves community who are mapping the S2R curriculum to their own objectives and standards. These efforts will ensure that the writing activities target skills valued within the classroom and professional writing community, making this interest-driven activity more broadly relevant.

Use rubrics. Sixteen projects used rubrics as an aid to score learner artifacts. Twelve projects developed rubrics for the assessment of specific artifacts, while four used general rubrics. Popham (1997) provided a succinct list of guidelines one should consider when creating and using rubrics. Generally, he points out that rubrics are difficult to write, and it is important that they are neither too broad nor too narrow. In the latter case, the rubric may constrain learning and hinder learner participation, a case of which is elaborated in the next main section.

---

[7] BuzzMath Q&A: http://www.hastac.org/dml-badges/buzzmath
[8] S2R Medals Q&A: http://www.hastac.org/dml-badges/S2R-Medals

S2R Medals had initially intended to use rubrics to assess both written and live reporting, but beta testing revealed the need to refine those rubrics so that both teachers and learners understood the assessment criteria. They also discovered that different rubrics were needed for different reporting settings; because live reporting is so complex, it can only be assessed by teachers who are present at the event.

Use formative functions of assessment. Fifteen projects provided varying types and amounts of formative feedback to learners. Five projects used peer feedback, three used expert feedback, and seven used a combination of the two. Schwartz & Arena (2009) make the case for choice-based assessments. Many researchers have argued that giving formative feedback enhances the learning experience (e.g. Black & Wiliam, 2009; Shepard, 2007), but Schwartz and Arena argue that the skill of knowing how to ask for formative feedback is a skill not being taught. Some projects encourage students to ask for, give, and use feedback to each other, which may help in building this skill.

The Providence After School Alliance's *Pathways for Lifelong Learning* [9] project provides Rhode Island youth with quality after school and extra-curricular activities. Like many projects, *Pathways for Lifelong Learning* aims to build a learning community within which learners can provide and receive critical feedback in safe and useful ways. In some activities, peers are expected to provide formative feedback in comments on one another's blog entries. Learners are also expected to participate on judging panels for the final demonstrations of activity artifacts and review the learner demonstrations with a rubric. These kinds of expectations aim to give learners a sense of agency and control over their own learning and assessment processes.

Enhance validity with expert judgment. Twelve projects used expert judges to evaluate learner artifacts. Nine used experts who were teachers or practitioners, two used computer scoring systems, and one project used an AI tutor. Popham's (2007) chapter on validity makes the complex concept of validity accessible by highlighting the information and practices teachers should consider to enhance the validity of the claims they make about learning. Messick's (1995) discussion of consequential validity is particularly important in a badge system context because the intended and unintended ways in which badges are used have a direct impact on learning outcomes and learner motivation.

---

[9] Pathways for Lifelong Learning Q&A: http://www.hastac.org/dml-badges/pathways-for-lifelong-learning

This was particularly salient in the *Design for America* (DFA)[10] project. DFA fosters innovation in a community where learners can share their designs and interests. Peer-to-peer interaction was an important aspect of the community, and peers were expected to provide valuable critical feedback to one another. This paired with expert feedback triangulated assessment in a way that made the badges more meaningful. Early feedback from users indicated that they wanted the badges to *do something* beyond recognizing an artifact. This challenge is described in the examples in the next main section. In this system, peer feedback is given and used for refinement purposes. Badges are not awarded because of feedback given by peers, but artifacts that earn badges may be influenced by that feedback.

Promote "hard" and "soft" skill sets. Eleven projects promote "soft skills" like leadership and collaboration in addition to the "hard skills" they promote. Schulz (2008) discussed the need for students to develop "soft skills" beyond academic knowledge. Many projects found that their students valued these "soft skills" that are transferrable to any domain over the "hard" content knowledge.

MOUSE Wins![11], a project that wants "the assessment process to be as social as the learning is," helps learners "[build] computational, digital, and workplace literacies" (HASATC Q&A). The project asserts that there is a feedback loop in the workplace; they want learning to mirror that organic process. They want to provide learners with the skills and literacies they need to successfully compete in the workplace, and they do that through teaching both hard and soft skills of the 21st century.

Use e-portfolios. Eight projects required learners to collect artifacts in a digital portfolio. One of these e-portfolio systems was open to the public, while seven were "closed," meaning only the immediate learning community could see and comment on them. Gillespie et al. (1996) provide a review of the recent literature on portfolio assessment and address the topic of private and public portfolios.

4H/USDA Robotics[12] entered this process with the intention of building upon their existing competition and portfolio system. In 4H, learners keep logs of their projects to track progress of each project, and turn those in for evaluation at the end of the project. In moving to a digital badge system, 4H/USDA Robotics aimed to engage learners in a networked community,

---

[10] Design for America Q&A: http://www.hastac.org/dml-badges/design-for-america
[11] MOUSE Wins! Q&A: http://www.hastac.org/dml-badges/mouse
[12] 4H/USDA Robotics Q&A: http://www.hastac.org/dml-badges/4-H-USDA-Robotics-Digital-Badges

within which they could submit artifacts for review by peers and experts, and receive more detailed feedback on their work.

Use performance assessments in relevant contexts. Seven projects used performance assessments to evaluate learners. Mehrens, Popham, & Ryan, (1998) provided six guidelines for using performance assessment, and suggested that instructors should be careful in how they prepare students for such assessments lest they compromise the assessment. The biggest challenge most projects found in utilizing performance assessments was making sure that the summative function of these assessments did not undermine the formative functions of the activity.

Sweetwater AQUAPONS[13] taught learners about biology through aquaponics. Learners kept a blog and photo diary documenting the growth and health of their aquaponics system and received feedback on their work through peer and expert comments. To earn a badge, learners had to demonstrate that their understanding of aquaponics systems through "written assessments, photo and video projects, and in-person demonstrations of proficiency" (Sweetwater AQUAPONS Proposal). However, the project kept the focus on the process over the product, thereby diminishing the potential deformative effects of the summative assessment.

Use mastery learning. In this context we use the term "mastery learning" to mean that learners are given practice until they have mastered a single skill set, and then move to the next skill set. Six projects did this, two of whom used humans to judge "mastery" and one used only a computer. Three projects used a combination of human and computer experts to judge mastery. Duncan & Hmelo-Silver (2009) define and discuss learning progressions, and advocate focusing on a smaller set of focused skills rather than a large set of skills in a perfunctory manner.

The Computer Science Student Network (CS2N) [14] comprised a host of informal and formal educational partners, including Carnegie Mellon's Robotics Academy and University of Pittsburgh's Learning Research and Development Center. They worked to provide scaffolded learning opportunities in science, mathematics, robotics, animation, and game design. In the CS2N project, badges in activities supported by AI tutors were validated through the AI tutor and through automated online testing (through Moodle), or automated detection of in-game events

[13] Sweetwater AQUAPONS Q&A: http://www.hastac.org/dml-badges/Sweet-Water-%28AQUAPONS%29
[14] CS2N Q&A: http://www.hastac.org/dml-badges/cs2n

(through Unity) in simulator environments. Instructor approval is used where appropriate in addition to automated tools.

Involve students at a granular level. Three projects involved students in the design of the physical badges, as well as in the design of the pathways it takes to earn a badge. Stefani (1994) studies student marks and grades, and their effectiveness in comparison to teacher marks. The projects that used this principle aimed to give learners agency over their own learning processes by including them in the design of activities and assessments. This practice turned out to be quite difficult, but projects learned a great deal and made significant changes to their system based on the feedback from the students.

Cooper-Hewitt's DesignPrep's[15] Badge Constellation Design Process intended to give learners the opportunity to design both the image of the badge and the pathway to the badge. As will be described in more detail below, the feedback they received directly affected a major shift in the kinds of activities and badges they awarded. Learners wanted to learn transferrable "soft skills" more than they wanted to learn content. Many of the learners had no interest in going into a design career, and wanted to use this opportunity to learn life skills.

Assessment Principles in Practice

This section presents three contrasting examples of how three principles were taken up differently in different contexts. These cases are meant to illustrate the important notion of *appropriate practice*. The principles outlined in this paper and their corresponding practice categories are not meant to be prescriptive; each assessment practice is dependent on its context, and some practices are better suited for certain contexts than others. Certainly new learning system developers can look to existing projects for guidance, but developers should keep in mind that the way assessment choices function within a particular learning system is wholly dependent on the specific context of the domain and community of learners. Some of the most exciting evidence that emerged as part of this research was the impact that different decisions about assessment practices had on learning and how this in turn connected with the research literature. The following examples illustrate these findings.

Use rubrics: Pathways to Global Competence and Design for America: Sixteen projects indicated that they intended to use rubrics to assess learning. Many projects asserted that using rubrics would make what was being assessed salient for both the learners and the assessors,

[15] Cooper-Hewitt DesignPrep Q&A: http://www.hastac.org/dml-badges/design-exchange

and they felt this transparency was important to earning the trust and support of the learners. Asia Society's *Pathways to Global Competence*[16] is a project aimed at helping youth build leadership skills. The project composed a series of very specific rubrics that were part of their Graduation Performance System. These rubrics aligned to their "Global Leadership Outcomes" and are intended to also eventually be aligned to the Common Core State Standards. These rubrics specify leadership characteristics, and the project maintains that this kind of specificity and transparency helps learners engage meaningfully in the activities.

Design for America also constructed specific rubrics for their learning activities, believing as did *Pathways to Global Competence* that transparency would allow learners to engage meaningfully. However, it seems that the project's focus on the broad topic of "innovation" attracted learners who wanted broader skills. When learners were given the rubrics, they reacted negatively, saying the activities felt "too much like school." The wanted the badges they earned and the activities around those badges to give them life skills; they wanted the badges to *do something*, not be constrained by learning goals in a rubric. *Design for America* took this feedback back to the development team and decided that, while they would not remove the rubrics, the rubrics would play a smaller role in the assessment process. The project decided to focus their efforts on skill building, collaboration, and leadership, valuing formative feedback and opportunities over summative assessments.

While rubrics do provide transparency, they also have the potential to constrain learning. Popham (1997) explains that rubrics must be carefully written so that they a neither too broad nor too narrow, because either extreme can have detrimental effects on learner engagement. If a rubric is too broad, it serves little purpose in defining the learning requirements. If it is too narrow, it can constrain learning and promote arguments over points that are not focused on the learning goals. These two projects enacted rubrics very differently given their different contexts, and their approaches have been successful. The lesson to be learned is that rubrics need to be tailored to the learning environment and community in which they will be used if they are to be effective.

Use formative peer feedback: Who Built America and Cooper-Hewitt DesignPrep. Twelve projects indicated that they would use peer feedback in some capacity to provide formative feedback to learners. Many projects accomplished this by creating mechanisms within

[16] Pathways to Global Competence Q&A: http://www.hastac.org/dml-badges/PASA

the learning community wherein learners could comment either directly on one another's artifacts, or provide critical feedback in a discussion forum. However, managing peer feedback is challenging, and attaching any kind of assessment to peer feedback proved to be problematic in some cases. The American Social History Project's *Who Built America*[17] project is a professional development site where history teachers can learn historical content and new curricular design skills. The highest badge in *Who Built America* is the Master History Teacher badge, which both indicates that the teacher is skilled in designing history lessons and allows the teachers to mentor and assess other teachers. The project intended to have learners comment on each other's lesson designs, and they wanted to award an *I <3 Collaboration* badge for collaborative efforts. The platform they were working within, however, could not support the assessment of collaboration, so the project had to think of a new way to foster this skill. The project ended up deciding to design learning activities that required participation in critical discussion forums. In this way, they encouraged collaboration, but avoided the messy issues around assessing a soft skill like collaboration by awarding an *I <3 Community* badge for participation in the forums instead of assessing collaborative efforts. In this way, the project may have ended up fostering more and more collegial collaboration than they would have in their original design.

Cooper-Hewitt DesignPrep (described above) also intended to engage learners in peer assessment, but they too found that formal peer assessment would be problematic. They decided to engage learners in activities that allowed them to provide critical feedback on learners' design artifacts. This severed to both improve learner-constructed artifacts and build community. Learners interacted with one another in both face-to-face and online settings, and the peer feedback component of this system helped to form productive relationships. While this feedback aided in the improvement of the final artifacts, the peer comments were not formally used to assess the final product. This practice served to build community and collegiality between learners in the face-to-face and online settings.

Both of these projects realized early on that a formalization of peer assessment may actually undermine the community building they were aiming to foster. Their assessment choices reflected their ability to balance the formative and summative functions of assessment without having deformative effects on learners or the community. Their formative feedback practices

[17] Who Built America Q&A: http://www.hastac.org/dml-badges/Who-Built-America

also served to engage learners in their own assessment process and have agency over their own learning.

Use e-portfolios: S2R Medals and SA&FS. Eight projects decided to use some kind of e-portfolio system, and all but one of those systems only allowed the learners in the community to see and comment on the artifacts. *S2R Medals* (described above) trains youth to become sports journalists. Earners' badges were associated with a portfolio of their work and encouraged not only the immediate learning community, but the community at large to comment on these portfolios and provide feedback. This meant that coaches, parents, and general community members were interacting with the learners and their projects, so learners received a variety of feedback from many different perspectives on their work. The varied feedback on the portfolios informed the revision of the artifacts, and served to demonstrate new and nuanced skills of the young journalists to the broader community.

In contrast, the *Sustainable Agriculture & Food Safety* (SA&FS)[18] program at UC Davis designed a digital badge system that augmented learners' classroom experiences with the opportunity to conduct a deep dive into a specific n area of interest in their field. The system allowed learners to design portfolios and present them to a board of judges for evaluation. These portfolios were viewable only to the immediate learning community. This closed system was in place partly for the privacy and protection of students, but it also facilitated more formal interaction around the portfolios. Learners decided what went into the portfolios and how they wanted to be assessed. This gave learners a sense of agency over their work, and allowed them to explore individual interests. They are an integral part of their own assessment, making the assessment process a challenging and transforming experience.

The advantages and disadvantages of public vs. private portfolios is an enduring strand of research in portfolio assessment (Gillespie, et al., 1996; Stiver, et al., 2011). One approach to portfolio assessment is not inherently better than the other, but the impact system developers' choices had on the kind of learning and revision in which the learners engaged is important. These two projects did not know about each other's practices. Part of the goal of this research is to connect projects and start a dialogue. When these two projects began conversing with one another, a very productive dialogue occurred, resulting in a better understanding of the kind of learning their assessments were fostering.

[18] SA&FS Q&A: http://www.hastac.org/dml-badges/SA%2526FS-Learner-Driven-Badges-Project

Implications of Assessment Choices for Learning Ecosystems

The Design Principles Documentation project began by following and documenting project practices, and connecting those practices to existing assessment literature. But it quickly became clear that an important function of this work was connecting projects with similar practices and engaging them in meaningful discussions about the thinking behind their design choices. By showing projects the practices they have in common with each other, productive dialogues have emerged and refinement of practices have occurred. This dialogue has also made projects explicitly aware of the impact their assessment choices have on learning, which they can then evaluate and refine as necessary. The design principles are instrumental in this dialogue, as existing projects can use them to find projects with whom they can engage, and new projects can review and ask questions to those already enacting these principles.

As education evolves more toward open and networked learning, innovations such as digital badges are becoming increasingly significant. For if one is going to use digital badges responsibly in education, they must consider the implications of their assessment practices on the learning process. By making the assessment design principles that emerged from the DML projects open to the public for use and discussion, this work is fostering important conversations about assessment design in digital credentialing and beyond.

By connecting these design principles to the scholarly literature and recommending assessment functions to consider when designing assessments, this work becomes relevant beyond creating a digital badge system; it is relevant to anyone designing assessments for educational programs. These assessment design principles offer a unique perspective on the implications of assessment design for learning, and can serve the larger audience as they design assessments within badging systems or in other contexts.

References

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability, 21(1), 5–31.

Brown, J. S, & Duguid, P. (2002). The social life of information. Harvard Business Press.

Brown, J. S., & Adler, R. P. (2008). Minds on fire: Open education, the long tail, and learning 2.0. Educause Review, 43(1), 16–20.

Brown, John Seely, Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. Educational Researcher, 18(1), 32–42.

Bulkley, K. E., Christman, J. B., Goertz, M. E., & Lawrence, N. R. (2010). Building With Benchmarks: The Role of the District in Philadelphia's Benchmark Assessment System. Part of a special issue: Benchmarks for Success? Interim Assessments as a Strategy for Educational Improvement, 85(2), 186–204. doi:10.1080/01619561003685346

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. Educational Researcher, 32(1), 9–13.

Darling-Hammond, L. (1997). Using standards and assessments to support student learning. Phi Delta Kappan, 79(3), 190.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. Journal of Research in Science Teaching, 46(6), 606–609. doi:10.1002/tea.20316

Gillespie, C. S., Ford, K. L., Gillespie, R. D., & Leavell, A. G. (1996). Portfolio assessment: Some questions, some answers, some recommendations. Journal of Adolescent & Adult Literacy, 39(6), 480–491.

Ito, M. (2005). Technologies of the childhood imagination: Yugioh, media mixes, and everyday cultural production.

Ito, M et al. (2009). Living and learning with new media: Summary of findings from the Digital Youth Project. Chicago IL: MacArthur Foundation.

Maxwell, J. A. (2006). Literature reviews of, and for, educational research: A commentary on Boote and Beile's" scholars before researchers." Educational Researcher, 35(9), 28–31.

Mehrens, W. A., Popham, W. J., & Ryan, J. M. (1998). How to prepare students for perforrnance assessments. Educational Measurement: Issues and Practice, 17(1), 18–22.

Oláh, L. N., Lawrence, N. R., & Riggan, M. (2010). Learning to learn from benchmark assessment data: How teachers analyze results. Peabody Journal of Education, 85(2), 226–245.

Popham, W. J. (1997). What's wrong-and what's right-with rubrics. Educational Leadership, 55, 72–75.

Popham, W. James. (2007). Classroom Assessment: What Teachers Need to Know (5th ed.). Allyn & Bacon.

Schulz, B. (2008). The importance of soft skills: Education beyond academic knowledge. Retrieved from http://ir.polytechnic.edu.na/jspui/handle/10628/39

Schwartz, D. L., & Arena, D. (2009). Choice-based assessments for the digital age. MacArthur 21st Century Learning and Assessment Project. Retrieved from http://dml2011.dmlhub.net/sites/dmlcentral/files/resource_files/ChoiceSchwartzArenaAUGUST232009.pdf

Shepard, L. A. (2007). Formative assessment: Caveat emptor. In C. A. Dwyer (Ed.), The future of assessment: Shaping teaching and learning (pp. 279–303). Mahwah, NJ: Erlbaum.

Stefani, L. a. J. (1994). Peer, self and tutor assessment: Relative reliabilities. Studies in Higher Education, 19(1), 69.

Stiver, W., An, J., Millen, B., Connors, R., & Sorensen, K. (2011). Development of an Engineering Portfolio System. Proceedings of the Canadian Engineering Education Association. Retrieved from http://library.queensu.ca/ojs/index.php/PCEEA/article/download/3729/3762

Yowell, C. M., & Smylie, M. A. (1999). Self-regulation in democratic communities. The Elementary School Journal, 469–490.

Table 1

Design Principles for Assessing Learning in Digital Badge Systems

General Design Principle (# of projects)	**Description of Principle**	**Specific Principles (# of projects)**	**Relevant Research Example**	**Example of a Specific Practice**
***Use leveled badge systems* (29)**	29 of the 30 projects included some kind of "leveling" system that students would move through as they practiced new skills, as opposed to a "flat" system where all badges have equal value.	Competency Levels (16) MetaBadges (10) Hierarchical Categories (3)	Bulkley et al. (2010) found that answers marked incorrectly did not help teachers in designing specific reteaching strategies.	Small badges for activities will add up to larger mastery badges (BuzzMath).
Align assessment activities to standards: Create measurable learning objectives (26)	*Standards varied from national and state standards to internal standards set by the parent organizations of the projects.*	Internal (10) National/State (7) Common Core (9)	Darling-Hammond (1997) discusses the need to raise standards and the system in which they are employed to support teaching and learning.	There is already a large community of teachers within the Make Waves community who are mapping the S2R curriculum to their own objectives and standards (S2R Medals).
Use rubrics (16)	Projects used rubrics as an aid to score learner artifacts.	Rubrics Developed for Assessment of Specific Artifacts (12) General Rubrics (4)	Popham (1997) provides a succinct list of guidelines one should consider when creating and using rubrics.	Rubrics are competency based and generated ad hoc by individual teachers. However, the project is looking to standardize the process and pull the rubrics into a system (LevelUp).
Use formative functions of assessment (15)	Projects provided varying types and amounts of formative feedback to learners.	Peer Feedback (5) Expert Feedback (3) Combination of Peer/Expert Feedback (7)	Schwartz & Arena (2009) make the case for choice-based assessments. Many researchers have argued that giving formative feedback enhances the learning experience (e.g. Black & Wiliam, 2009; Shepard, 2007) but Schwartz and Arena argue that the skill of knowing how to ask for formative feedback is a skill not being taught.	At high school level peers participate as panel judges for the final demonstrations and review the student demonstration with a rubric. Peers are also expected to provide formative assessment on peer blog entries online (Pathways for Lifelong Learning)

Table 1 (Continued)

General Design Principle (# of projects)	Description of Principle	Specific Principles (# of projects)	Relevant Research Example	Example of a Specific Practice
Enhance validity with expert judgment (12)	Projects used expert judges to evaluate learner artifacts.	Teacher/Practitioner Experts (9) Computer Scoring System (2) AI Tutors (1)	Popham's (2007) chapter on validity highlights the information and practices teachers should consider to enhance the validity of the claims they make about learning.	Badges are validated by community mentors. Peer feedback is given and used for refinement purposes. Badges are not awarded specifically because of feedback given by peers, but the artifacts that earn badges may be influenced by that feedback (Design for America).
***Promote "hard" and "soft" skill sets* (11)**	Projects promote “soft skills” like leadership and collaboration in addition to the “hard skills” they promote.	N/A	Schulz (2008) discusses the outcry for students to develop “soft skills” beyond academic knowledge.	The project wants "the assessment process to be as social as the learning is." There is a feedback loop in the workplace, and they want learning to mirror that organic process (MOUSE Wins!).
Use e-portfolios (8)	Projects required learners to collect artifacts in a digital portfolio.	Open to Public (1) Local to Community (7)	Gillespie et al. (1996) provide a review of the recent literature on portfolio assessment and address the topic of private and public portfolios.	“Every S2R participant has their personal Reporter Page on www.makewav.es/s2r This serves as an e-portfolio and permits their educators, supporters, friends, family and peers to see and evaluate their work” (S2R Medals).
Use performance assessments in relevant contexts (7)	Projects used performance assessments to evaluate learners.	N/A	Mehrens, Popham, & Ryan, (1998) provide six guidelines for using performance assessment, and suggest that instructors should be careful in how they prepare students for such assessments lest they compromise the assessment.	“The badges for each curricular area will be earned through written assessments, photo and video projects, and in-person demonstrations of proficiency” (Sweetwater AQUAPONS).

Table 1 (Continued)

General Design Principle (# of projects)	Description of Principle	Specific Principles (# of projects)	Relevant Research Example	Example of a Specific Practice
Use mastery learning (6)	In this context we use the term "mastery learning" to mean that learners are given practice until they have mastered a single skill set, and then move to the next skill set.	Judged by Human Experts (2) Judged by Computer Experts (1) Combination of Human and Computer Experts (3)	Duncan & Hmelo-Silver (2009) define and discuss learning progressions, and advocate focusing on a smaller set of focused skills rather than a large set of skills in a perfunctory manner.	Badges in activities supported by AI tutors are validated through the AI tutor and through automated online testing (through Moodle), or automated detection of in-game events (through Unity) in simulation environments. Instructor approval is used where appropriate in addition to automated tools (CS2N).
Involve students at a granular level (3)	*Projects involved students in the design of the physical badges, as well as in the design of the pathways one must take to earn a badge.*	N/A	Stefani (1994) studies student marks and grades, and their effectiveness in comparison to teacher marks.	In the Badge Constellation Design Process, Cooper-Hewitt is realizing that the badges should have "personality" and personal touched added by students. The process of designing a badge reflects the process that goes into receiving a badge (Design Exchange).