What are the learning and assessment objectives in educational Fab Labs and Makerspaces?

Yoav Bergner
New York University
yoav.bergner@nyu.edu

Samuel Abramovich
University at Buffalo
samuelab@buffalo.edu

Marcelo Worsley
Northwestern University
marcelo.worsley@northwestern.edu

Ofer Chen
New York University
ofer.chen@nyu.edu

ABSTRACT

We report on the analysis of responses about the benefits of educational Fab Labs and Makerspaces (FLMs) and about claims of student growth and learning in these spaces. We review related literature on assessment in FLMs and on efforts to develop frameworks. Methods of the the current study include the design of the FabLearn workshop, where responses were elicited from practitioners, and mixed-methods analyses of those responses. Results, in the form of the coding guides, are presented, followed by a discussion of the construct categories, alignment between benefits and claims codes, and consequences for coherent assessment in FLMs.

KEYWORDS

Assessment, Maker Education

ACM Reference Format:
Yoav Bergner, Samuel Abramovich, Marcelo Worsley, and Ofer Chen. 2019. What are the learning and assessment objectives in educational Fab Labs and Makerspaces?. In *FabLearn 2019 (FL2019), March 9–10, 2019, New York, NY, USA.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3311890.3311896

1 INTRODUCTION

There has always been an interdependent relationship between formal assessment and instructional practices in education. Assessment traditionally serves to provide accountability for educational outcomes, as in the case of tests used for summative purposes. But assessments, whether used for formative or summative purposes, can also focus the efforts of instruction, as noted in the Next Generation Science Standards [20] or Standards for Technological Literacy [11]. In general, the practice of articulating claims about student knowledge, skills, and abilities/attitudes/aptitudes (KSAs) brings what is "counted" in the classroom [7] into relief. A healthy educational ecosystem may depend on coherence between assessment and instructional practice, and this coherence likely takes a sustained effort to achieve [30]. Assessment cannot serve students well if teachers, parents, or other stakeholders have little faith in its connection to instruction. In the context of educational Fab Labs and makerspaces (FLMs), it thus makes sense that the Agency *by* Design [5] study group has urged policy makers to support efforts to document and assess maker-centered learning.

With an eye toward developing an assessment framework that is coherent with instructor values and practices, we convened a practitioner workshop on assessment of learning in FLMs in 2016. Our goal was to bring assessment specialists and instructor-practitioners into a structured discussion about assessment issues in these learning spaces. Values and practices would be contributed by practitioners, but the language of claims and evidence used in the discussion would be shaped in accordance with the framework of evidence-centered assessment design (ECD) [19]. ECD urges one to work backwards from the claims one wants to be able to make about students (e.g., KSAs) to an articulation of the types of evidence that would support those claims (the assessment "argument") and then finally to the types of tasks that provide opportunities to elicit the desired evidence. The questions raised in the workshop were oriented around the first part of this process (i.e., on the claims). However, the workshop also sought to situate those claims in the context of how educators think about the value of FLMs more broadly and to identify current practices and perceived obstacles.

This paper focuses on the analysis of responses collected in part of the workshop, specifically responses to prompts about the benefits of FLMs and about claims of student growth and learning in these spaces. We first briefly review the related

FL2019, March 9–10, 2019, New York, NY, USA

ACM ISBN 978-1-4503-6244-3/19/03…$15.00
https://doi.org/10.1145/3311890.3311896

literature on assessment in FLMs and on efforts to develop assessment frameworks. We then describe the methods of the current study, including the design of the workshop and the mixed-methods analysis of responses collected from practitioners. The results, in the form of the coding guides, are presented and followed by a discussion of the emergent construct categories and consequences for future work.

2 BACKGROUND ON ASSESSMENT IN FLMS

It is worth noting that discussions of assessment in FLMs have been commonly absent until recently. Many articles on the potential of and state-of-the art of educational FLMs make no mention of assessment at all (e.g. [2, 9, 17, 23, 27]). Martinez and Stager's [18] sole mention of assessment was to point out that it "always interrupts the learning process"[18, p.81], that is to grudgingly acknowledge the existence of assessment and to object to it. Blikstein and Worsley [4] point to other examples of a "dismissive stance" within the Maker Movement to assessment of student learning. In the publication of the two-volume edited collection Makeology [22], a handful of chapters did engage with issues of assessment to some degree with Wardrip and Brahms [29] identifying assessment-related challenges for bringing making into schools, while Fields and Lee [14] described a rubric for assessment of crafting projects. Martin and Barron [16] approached the issue of assessing digital media citizenship head on, including articulation of a conceptual framework and the use of survey instruments. They chose to focus on noncognitive dispositions towards constructive production, social advocacy, and critical consumption. Blikstein, Kabayadondo, Martin, and Fields [3], by contrast, developed an assessment instrument for technological literacy. The authors identified only one comparable prior example for assessing student engineering design in a maker-oriented context. By their own admission, their instrument may be more technocentric [21] than what some maker educators might desire.

We are aware of a few holistic frameworks for assessment in FLMs, with many of them works in progress. Before describing some of these in detail, we wish to highlight a distinction between construct frameworks and types of assessment data (e.g., surveys, observations, rubrics). Data types, as categories, can answer questions like, "what kinds of evidence are collected?" or "How do you assess?" Construct frameworks answer the thornier questions, "Evidence for what? What do you assess?" The MakerEd Open Portfolio Project [6], for example, focuses attention on issues related to portfolios as a data type. That an educator might use student portfolios leaves open the question of whether one uses them to assess creativity or technical proficiency or both. The focus of this paper is on construct frameworks and the question of "what." When a holistic construct framework defines categories for the evidence-for-what question, it is logical to then expect the basis on which these categories were selected over other categories. The frameworks discussed here involved a process similar to Agency *by* Design's [5] report, which synthesized the collection of data from practitioners (interviews and site visits) with extant literature on maker-centered learning outcomes.

The A*b*D report stated that "the most important benefits of maker education are... developing a sense of self and a sense of community" [5, p. 7]. Empowerment is thus at the core of A*b*D's framework. It is not clear how empowerment would manifest itself until it is operationalized as sensitivity to design. Sensitivity to design is further subdivided into categories, which do span observable "moves" or "indicators": looking closely, exploring complexity, and finding opportunity. Each of these categories comprises roughly five different practices, and the list of practices can, for example, be integrated into a classroom observation protocol.

Wardrip and Brahms [28] also put learning practices forward in their framework, which is specifically designed for informal learning spaces such as museums. Perhaps because of the focus on informal spaces where learner experiences are typically short (an afternoon), the authors make no specific claims about psychological constructs. Their list of practices spans: Inquire, Tinker, Seek Share Resources, Hack & Repurpose, Express Intention, Develop Fluency, and Simplify to Complexify. Within observational protocols, each of these practices is further refined. For example, "seeking" can apply to tools, information, or other supports.

While the prior examples described are frameworks centered on practices, two other contemporary frameworks do point to psychological constructs that underlie observable practices. Beyond Rubrics (BR) is a research project that emphasizes embedded (and often playful) assessments and co-design processes for makerspaces. BR is construct-centered, but it is not clear from currently published materials how the set of constructs was finalized. The key constructs, referred to as "Maker Elements," are Agency; Design Process; Social Scaffolding; Productive Risk-Taking; Troubleshooting; Bridging Knowledge; and Content Knowledge. Troubleshooting may seem like a practice rather than a construct; however, BR materials describe Troubleshooting using the language of skills, traits (in the voice of the learner): "I have the skills, tools, and persistence to solve problems. . . " It is worth noting that, especially as worded, troubleshooting "skill" could be confounded with learner agency. It might also be confounded with design process, which is summarized by "I can plan, create, test, and iterate my designs."

As a final example, the MakEval [15] project identified five key "targets," which combine both constructs and practices: Agency, STEM Practices, Creativity, STEM Interest and Identity, Critical Thinking. The basis for the categorization

Table 1: Some holistic frameworks for assessment in FLMs and their categories.

Framework	Key Constructs or Practices
Agency *by* Design [5]	Sensitivity to design: looking closely, exploring complexity, and finding opportunity
Wardrip and Brahms [28]	Inquire, Tinker, Seek & Share Resources, Hack & Repurpose, Express Intention, Develop Fluency, and Simplify to Complexify
Beyond Rubrics [13]	Agency; Design Process; Social Scaffolding; Productive Risk-Taking; Troubleshooting; Bridging Knowledge; Content Knowledge
MakEval [15]	Agency, STEM Practices, Creativity, STEM Interest and Identity, Critical Thinking

is "formal and informal maker educators' survey and interview data," although specific methodological details have not yet been published. MakEval aims to provide educators with suites of tools, such as surveys, rubrics, and observation protocols, for each of the different target areas. The key constructs and practices from the four holistic frameworks discussed above are collected in Table 1.

In sum, we have pointed to some specific examples of targeted assessments, for example of craft production, technological literacies, and citizenship-oriented dispositions. We have also described some more holistic frameworks and identified some of the variance between them. Differences include both broad focus (psychological constructs vs. practices) and specific identification of practices or constructs. Of the example frameworks described above, we were not always able to identify details of the methodology used in the aggregation of data from survey participants. Consequently, this manuscript is partly a substantive contribution and partly a methodological one. It is substantive in the sense that it attempts to identify dimensions for holistic assessment in FLMs. Readers may use our results as a springboard for specific assessment planning. But our work may also be a methodological contribution in that we focus attention on how we arrived at the categories presented in the results. Readers may follow our process to create their own categories. The mixed-methods approach we describe here is certainly not the only way to aggregate knowledge from a mix of instructors and assessment specialists, but we hope it may at least serve as a model for discussion of alternatives.

3 METHODS

Data Collection and Participants

Data for this analysis were collected before and during a practitioner workshop coordinated by a subset of the authors. The purposes of the workshop were (a) to build community between instructors and assessment specialists around assessment of learning in FLMs and (b) to gather data from a structured conversation about growth claims. Before the workshop, an electronic survey was sent out via email to registered participants. As an incentive to participate, responses to the surveys counted as an entry ticket to a lottery for $50 gift card. The pre-workshop survey collected descriptive data from the participants including (a) how much experience they had in making, (b) the structure and context of their experience, and (c) their role.

Participants were asked to respond to a small number of Likert-type items, for example: "Generally speaking, how do you feel about the importance of assessment?" [Five point scale: I couldn't care less (1), Somewhat Important (2), Important (3), Very Important (4), I dream about assessment (5)]. There were also several open-ended questions about the benefits of making/makerspaces, limitations and challenges to success, the usefulness of assessment, and types of evidence to demonstrate the benefits of makerspaces. From the pre-workshop survey, we concentrate here on responses to the first of the open-ended questions: P1-Benefits: What are the three (to five) biggest benefits provided by Making/Makerspaces? The workshop itself was conducted in three parts, with short presentations at the beginning of each part. The presentations were designed to promote an expansive way of thinking about assessment and did not promote particular psychological constructs or practices. After each presentation, participants worked in small groups of 4-5 to discuss one question, and at the end of that discussion, each participant took time to record their own notes in response to the question. These notes were collected for analysis. We note that while responses were collected from individuals, they were not independent samples, since the individuals were in discussion with their group partners. As a result, certain responses were likely to be multiplied in cases where group members agreed. Nevertheless, there were also differences in the way that individuals articulated the ideas of the group. As a collection methodology, having individuals work in groups was a compromise between having independent samples and the benefit of socially negotiated understanding. Responses to P1-Benefits, in the pre-workshop survey, were independent.

The three questions for workshop discussion were: Q1-Claims: What should a Fablab assessment tell people about the learner? Q2-Process: What is the most valuable process/method for you to determine if someone learns in a Fablab? Q3-Challenges: What do you think are the biggest challenges for assessments in Fablabs?

Consistent with our focus in this analysis on the constructs of interests (evidence-for-what) and not on specific methods of evidence identification, we restrict our attention in the remainder of the manuscript to the responses from P1-Benefits, administered individually over email pre-workshop, and Q1-Claims, administered after group discussion during the workshop.

Data Processing

Open-question survey responses collected during the workshop (handwritten) were entered into a spreadsheet for further processing. The responses, provided as free text, were parsed to more granular response-items, which allowed for coding of the sub-parts under different constructs. For example, in response to Q1-Claims, one participant answered: "I would like to know if the learner comes away with a better sense of their STEM identity, How the making experience drew them closer to feeling like they are creative and can solve problems, Do these experiences make them [more] likely to persist in school environments"

This response was parsed to the following items: If the learner comes away with a better sense of their STEM identity; How the making experience drew them closer to feeling like they are creative; [How the making experience drew them closer to feeling like they] can solve problems; Do these experiences make them [more] likely to persist in school environments?

In Q1-Claims, the assessment unit of analysis is the student, and therefore classroom, school, or instruction-level claims were removed from the analysis. For example, one response included: "...The overall view of the classroom, Capacity + leaning of the facilitator/PD overall/gaps in PD. . . "

It is clear that in this case, the participant had suggested assessments of the classroom and of the instructor rather than the student. The parsed items from this example were therefore removed from the item pool. Additional items were removed when the response was a strategy or method of assessment, that is, a suggestion for "how" we should assess students instead of "what" should be assessed about students. Examples included "certification" or "video capture." Finally, a small number (3 items from P1-Benefits and 14 from Q1-Claims) of parsed items were indecipherable and therefore removed. After conducting these procedures the final item pools for P1-Benefits and Q1-Claims consisted of 111 and 284 items respectively. Results are summarized in Table 2.

Table 2: Summary of the data processing phase.

	P1-Benefits	Q1-Claims
Initial responses	35	49
Initial pool of parsed items	114	323
Not student-level	0	13
Strategy for assessment	0	12
Indecipherable items	3	14
Final item pool	111	284

Mixed-methods Analysis

Mixed methods conversion studies refer to the process of converting one data type into another for the purpose of analysis. Conversion studies include qualitizing quantitative data so that they can be analyzed qualitatively, or quantitizing qualitative data for the purpose of statistical analysis, or both [26]. Bazeley [1] emphasized the possible benefits of conversion designs, at times including multiple data conversions, in enhancing inference making in mixed methods research.

The following conversion process was carried out separately in its entirety for P1-Benefits and Q1-Claims.

Open coding. Initially, parsed responses were analyzed qualitatively by three of the authors of this paper. Each researcher open-coded all items using the method of constant comparisons [10]. This resulted in three sets of codes, one set for each coder. This process of open coding, when done individually by multiple coders, can lead to the creation of very different categories. Though different, each of these sets of codes can have intuitive meanings, maintain coherence, and yield useful analytic results [24]. This case was no different and revealed large differences in the ways the coders deemed appropriate to categorize the parsed items.

Quantitizing and factor analysis. To reconcile the differences between sets, and with the goal of converging on a single set, the next phase involved quantitization and exploratory "factor analysis." We put the term in quotes to emphasize that factor analytic methods were used here to identify unifying structure in the category schemes used by the three coders, not structure "inherent" in the items themselves. First, the number of factors was selected. The qualitative data (coded items) were quantitized by assigning binary values (dummy-coding) to the categories, combined from all coders. That is, for each item (row), a "1" was assigned to the categories (columns) that were used by the coders and a "0" was assigned to remaining categories. These item-category binary tables were uploaded to R for exploratory factor analysis (EFA). The number of factors to keep was to be decided using parallel analysis (PA) [25], but eigenvalues of the tetrachoric correlation matrix were also examined to check whether PA led to a result consistent with the eigenvalues-greater-than-1 rule. After determining the number of factors, factor loadings for the individual coding categories were produced using a varimax rotation. Loadings below a threshold value of 0.4 were ignored. The labels from categories that loaded onto common factors were then combined for the next stage, which focused on making sense of the reduced set of categories and the creation of a coding guide.

Qualitizing the factors and creating an initial coding guide. After categories created by individual coders were combined based on the factor analysis, the content of the new resultant categories was examined. In some cases, upon closer inspection, a "category" was revealed to be an "other" or catchall bucket for difficult-to-code responses. This factor was deleted, on the understanding that a coding guide would either clarify what to do with these responses, or they would be deemed uncategorizable. In other cases, two categories that had significant overlap could be combined, after it was resolved that coders might have used fine distinctions that discriminated only between some of the cases. It was understood that the coding guide would help to clarify multiple potential interpretations of the category label. On the basis of this examination, an initial coding guide was developed.

Validating the coding guides. Two more rounds of coding then took place in order to detect ambiguities and refine the coding guide. Each round used non-overlapping subsets of the response-items (roughly one-third to one-half), such that the researchers did not repeatedly code the same items. In each round, items were coded according to the newly created guides by the authors. Any discrepancies were discussed and clearer boundaries for each category were negotiated. Agreement was determined using Krippendorf's α [12], which reduces to Fleiss' κ [8] if all raters code the same nominal data set, and a threshold value of 0.7 was considered acceptable. (This level was achieved in one round for the P1-Benefits item pool and in two rounds for Q1-Claims.) When satisfactory agreement by the researchers was reached, the final coding guides were used to train two graduate students, who were not previously involved in the process. Training involved coding a sample of 10-15% of the response-items, after which the trainees answers were compared to each other and to expected answers (all training items had agreement by at least three out of four of the researchers). Following this, the trainees independently coded the remaining items in P1-Benefits and one-third of the remaining data set for Q1-Claims. Inter-rater agreement and the final coding guides will be reported in the results section.

4 RESULTS

For clarity, we present some of the results in each stage of the conversion study for P1-Benefits, but we provide only a summary of the results for Q1-Claims.

P1-Benefits

The open coding process for P1-Benefits generated three sets of categories. One coder used six categories, another used nine, and the third used ten categories. Both parallel analysis and eigenvalues of the tetrachoric correlation matrix pointed to nine underlying factors. For each factor, between two and four categories had a loading value of over 0.4, which provided a basis for combining the categories from different coders. One factor had no clear characteristics or attributes and was essentially a combination of "other" categories used by the individual coders. It was decided that most of the response-items in this factor could easily fit in one of the other eight categories.

As an example of categories from the first round of P1-Benefits that could be combined, the following all loaded strongly on one of the factors: 'Group Engagement', 'People Resources', 'Collaboration', and 'Collaboration and Community Orientation'. The emergent category captured the perceived benefits of teamwork, collaboration, and productive use of human resources. The remaining eight factors are summarized in the coding guide, Table 3.

The coding guide was tested using two previously untrained graduate students, as described in section 3.3.4. After the training phase, the two raters achieved an agreement (α = κ in this case) of 0.75, which was satisfactory.

Q1-Claims

The initial factor analysis for Q1-Claims resulted in 11 factors, of which one was deemed an "other" category and deleted. During the subsequent qualitizing phase, three of the remaining factors were reduced to two. These three factors appeared to span some combination of "mindset" constructs, confidence, grit, persistence, personal goals, meaning, and interest. It made sense that there would be correlation among these expressions of student traits and growth, but it seemed possible to organize them among two factors rather than three, one bucket for "emotional resilience" and another for personal values, interest, and identity. See the coding guide for assessment of the learner in Table 4. As for P1-Benefits, the coding guide was tested using two previously untrained graduate students, who achieved an agreement ($\alpha = \kappa$) of 0.73 after training. In Section 5, we discuss some of the similarities and differences between the final coding guides for P1-Benefits and Q1-Claims and connect these the constructs to related literature.

5 DISCUSSION

In the previous section, we presented results from separate analyses of the terms that maker educators use when describing (a) benefits associated with Making and FLMs and (b) claims about learner growth in these spaces. Each analysis produced a set of eight codes, shown in Tables 3 and 4. The code sets, while not exactly congruent, do overlap significantly. We therefore explore the mapping between these two sets of codes in the first part of this section. We then

Table 3: Benefits of Making/Makerspaces coding guide.

Category	Definition	Short exemplars
PS - Problem Solving Mentality	The acquisition of skills and ways of thought that an engineer, for example, would possess but that are not directly tied to specific tools or technologies	Critical thinking, problem solving, ingenuity, innovation, systems thinking
TM - Teamwork Experience and Skill Acquisition	Refers both to development of skills beneficial to working in a team (communication, collaboration) and to the affordances of teamwork (guidance and mentoring opportunities)	Collaboration, communication, reciprocal teaching, mentoring, interdependence, complementarity
HO - Hands-on Learning	Learning in activity which leads to an observable physical/behavioral engagement (but not necessarily to affective/psychological engagement)	learning in activity/by doing, crafting, making, active, (observable) engagement
BCW - Breaching the Classroom Walls	Enabling students to form a meaningful connection between the learning in the makerspace to solving real world problems. Mimicking the type of work and learning that later occurs in the real world through a more authentic experience	Real life relevance, open-ended project-based learning, unique experience, interdisciplinary
ER - Emotional Resilience	The development of emotional fortitude that increases challenge and learning seeking behaviors (agency, accountability) and enables the learner to overcome those challenges (persistence, failure management)	Empowerment, grit, persistence, failure management, confidence, agency, "mindset", accountability for own learning, self-efficacy
IN - Inclusivity	Providing learners from diverse backgrounds the opportunity to participate and be challenged in working on STEAM projects, along with an exposure to a variety of tools and concepts that may not be accessible otherwise	Diversity, access to STE(A)M, exposure
TTL - Technical and Technological Literacy	The acquisition of knowledge and skills that allow the proper use of tools and technology and support future learning of new tools.	Fabrication skill building, tool acquisition, high tech software and equipment
SE - Self Expression	The ability to express one's interests, values, and imagination through the creation of a personally meaningful object. Bringing 'yourself' to the classroom.	Interest building, self-directed, creative, discovery learning

Table 4: Claims about the learner coding guide.

Category	Definition	Short exemplars
DT - Design Thinking	Assessment of skills associated with proficient designers or problem solvers (but not specific technical skills or literacies)	Design, critical thinking, problem solving, iteration, prototyping
RPM - Reflection, Process, and Metacognition	Assessment of the learner's self-regulation and metacognition in understanding, and documenting a productive design process	Process, reflection, self-regulation, organization, articulation, evaluation, planning
CC - Collaboration and communication skills	Assessment of skills associated with productive teamwork. The degree to which a learner can be a contributing member of a team or community.	Collaboration, communication, community (of practice)
TS - Technical and fabrication skills	Assessment of technical skill, technological literacy, or fabrication competencies	Fabrication skill building, tool acquisition, high tech software and equipment
FC - Flexibility and Creativity	Assessment of learner's ability to view/solve a problem in more than one way, to try out different approaches, and to adapt to unexpected or novel situations	Creative and divergent thinking, innovative risk taking, cognitive flexibility
ER - Emotional resilience	Assessment of the emotional fortitude, self-directed learning, challenge-seeking behaviors, and persistence in face of challenges	Empowerment, grit, persistence, failure management, confidence, mindset, accountability for own learning, self-efficacy
PI - Personal interests, values, and identity	Assessment of the factors that mediate learning through motivation. What the learner brings with them and their affinities	Personal goals, intrinsic motivation, interest, agency, fears, communitarianism
GCG - General claims about growth		Progress, growth curves, what they learned, transfer

relate our findings to the constructs and practices identified in prior literature.

Mapping P1-Benefits to Q1-Claims

What can we learn from the ways in which the code sets for benefits and claims align and differ? Even before examining the codes themselves, one might anticipate the following logic in the mapping between benefits and growth claims. All growth claims about learners in FLMs imply (map to) benefits of these spaces, but not all benefits take the form of growth claims at the level of the learner. Benefits that do not map to claims here may be understood to apply at a larger systemic level than the individual, which is the target in our Q1-Claims phrasing ("about the learner"). We might call this source of mis-alignment a structural difference between the existing sets of benefits and assessment claim. Alternatively, benefits may fail to map to claims due to conceptual differences; that is if maker educators think about benefits and assessment claims in essentially different ways. Since the codes were products of an interpretative process, conceptual differences might also manifest in mappings between benefits and claims that are not one-to-one.

Indeed we find many of these expectations confirmed.

For example, technical skills and technological literacy emerged in both perceived benefits and construct-oriented claims about FLMs using consistent language. The same can be said about Emotional Resilience—which is perhaps less obvious—as well as Teamwork and Collaboration. Even if the benefit and claims language differs a bit regarding teamwork, the usage differences might be resolved, for example, by noting that teamwork experience is a perceived benefit and can lead to the development of collaboration skills.

Some codes from the benefit and claim sets align but with some admixture between labels. For example, self-expression relates to personal interest and identity, which is also grouped with values. Looking at claims, however, we have a category labeled Flexibility and Creativity. The word "creativity" is sometimes used with a connotation of ingenuity and innovation (that is, an entrepreneurial sense), but it is also used to refer to creative self-expression. Thus, there may be a need to clarify the boundaries in the presence of semantic ambiguities. Cognitive flexibility is distinct from other aspects of Design Thinking, but both may be associated with a Problem Solving Mentality and the ability to ideate effectively when faced with new or challenging problems.

It is less clear how Reflection, Process & Metacognition (RPM), from the claims codes, should map to the benefits. Although somewhat related to Problem Solving , RPM also refers to more general practices and constructs often associated with learning how to learn. The fact that this category did not emerge from our analysis of the P1-Benefits responses could be an accident of chance, or it could point to conceptual differences in thinking. For example, maker educators may be aware that they want learners to be reflective of their making process, but growth along this dimension does not occur to them as a particular benefit of FLMs. The general claims about growth and transfer (GCG) category emerged in response to prompts regarding claims but not benefits. Responses in this category were vague (see Table 4) and might be coming from a place of discomfort around accountability.

Finally, there were three benefit codes—Hands-On Learning, Breaching the Classroom Walls, and Inclusivity—that did not appear to correspond directly to claim codes. Inclusivity (IN) is a benefit associated with diversity and access. While inclusivity affects individuals, it would not be assessed at the level of the individual. Rather, it is a systemic property and thus represents a structural difference in the focus (individual vs. collective).

Breaching the Classroom Walls (BCW) is similarly systemic, in that interdisciplinary and authentic practices can occur (or not) in FLMs. However, in contrast to inclusivity, it is less clear why this should be considered a positive end-in-itself. To put it another way, it would not be inappropriate to ask what learner skills, abilities, or attitudes would be changed for the better by experiences that breach the classroom walls. In the case of teamwork experience, for example, the intended effect was improved collaboration and communication skill. Exploring this question for BCW is an opportunity to refine an assessment framework with implications for instructional design. The same might be said for Hands-On Learning (HO). BCW and HO are identifiable practices in FLMs, but, at least in the present study, neither is clearly linked to growth and constructs.

Directions for future work

Taking a step back, the categories identified for benefits and claims encompass many of the constructs and practices mentioned in prior literature. There is no shortage of discussion of the design process, creativity, and critical thinking within the FLM literature. Common perceptions about importance of agency and STEM identity development are also paralleled in the categories of Self Expression (SE) and Personal interests, values and identity (PI). However, there were two categories that emerged within this data set but were less prominent in other frameworks. Both Emotional Resilience (ER) and Teamwork/Collaboration Skills (TM/CC) were robust construct-oriented categories, with clear alignment in responses to both perceived benefits and claims questions. Of the holistic frameworks reviewed in Table 1, Beyond Rubrics comes close to these categories through the practices of productive risk-taking and social scaffolding. However, those practices are also much more narrowly defined. Although beyond the scope of this paper, there has been a great deal of thought given to collaboration and emotional resilience in educational contexts outside of FLMs.

If a healthy FLM ecosystem depends on a coherence between assessment and instructional practice, then the results summarized in Tables 3-4 suggest some starting points both for assessment and instruction. Developing assessments of emotional resilience and collaboration skill may be challenging, but it is important to pursue advances in these directions in order to optimize benefits to learners in FLMs. At the same time, instructional design can be informed by these growth goals. Perhaps breaching the classroom walls should not be seen as an end in itself but rather linked to cultivation of new interests and values. Hands-on learning, valued because of its association with sustained engagement, might be seen as a means to end in the cultivation of emotional resilience or of metacognition. Importantly, a clear framework can also help practitioners reject the wrong kind of assessments. For example, the use of assessments borrowed from other learning contexts (e.g., a standardized test or group-work portfolio rubric) can be a source of incoherence if the connection to the specific learning context is not clear. Regarding the sustained effort necessary to achieve coherence, the benefits and learning objectives identified here were common to the vast majority of the participants in the workshop. This suggests that an assessment for FLMs that is based on those elements can also rely on sustained effort from FLM facilitators.

Finally, we hope that both our methodology and findings can serve as part of the necessary cohort of research that will lead to more FLM appropriate assessments that are based on construct frameworks and not strictly defined by available data. Even if only our methodology is borrowed by other assessment developers, we expect the results will address what is evidence used for and how it can help the learner.

ACKNOWLEDGMENTS

The authors would like to thank all of the participants in the workshop where data for these analyses were collected.

REFERENCES

[1] Pat Bazeley. 2006. The contribution of computer software to integrating qualitative and quantitative data and analyses. *Research in the Schools* 13, 1 (2006), 64–74.

[2] Paulo Blikstein. 2013. Digital Fabrication and 'Making' in Education: The Democratization of Invention. In *FabLabs: Of Machines, Makers and Inventors*, J. WalterHermann and C. Boching (Eds.). Transcript Publishers, Bielefeld, Chapter 1, 1–21.

[3] Paulo Blikstein, Zaza Kabayadondo, Andrew Martin, and Deborah Fields. 2017. An Assessment Instrument of Technological Literacies in Makerspaces and FabLabs. *Journal of Engineering Education* 106, 1 (2017), 149–175.

[4] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal Learning Analytics and Education Data Mining: Using Computational Technologies to Measure Complex Learning Tasks. *Journal of Learning Analytics* 3, 2 (2016), 220–238.

[5] Agency by Design. 2015. Maker-centered learning and the development of self: Preliminary findings of the Agency by Design project.

[6] S Chang, A Keune, K Peppler, A Maltese, C McKay, and L Regalla. 2015. Open portfolio project: Research brief series.

[7] National Research Council. 2001. *Knowing what students know: The science and design of educational assessment.* National Academies Press.

[8] Joseph L Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76, 5 (1971), 378–382.

[9] Neil Gershenfeld. 2008. *Fab: the coming revolution on your desktop–from personal computers to personal fabrication.* Basic Books.

[10] Barney G. Glaser and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research.* Sociology Press, Mill Valley, CA.

[11] International Technology Education Association. 2001. *Standards for Technological Literacy: Content for the Study of Technology* (3rd editio ed.). Vol. 82. 513–517 pages.

[12] Klaus Krippendorff. 1970. Estimating the reliability, systematic error and random error of interval data. *Educational and Psychological Measurement* 30, 1 (1970), 61–70.

[13] Teaching Systems Lab. 2018. Beyond Rubrics: Assessment in Making. https://tsl.mit.edu/projects/beyond-rubrics/

[14] Victor R Lee and Deborah A Fields. 2017. A rubric for describing competences in the areas of circuitry, computation, and crafting after a course using e-textiles. *International Journal of Information and Learning Technology* 34, 5 (2017), 372–384.

[15] Adam V. Maltese. 2018. MakEval: tools to evaluate maker programs with youth. http://www.adammaltese.com/content/makeval/

[16] Caitlin K Martin and Brigid Barron. 2016. Making matters: A framework for assessing digital media citizenship. In *Makeology*. Routledge, 59–85.

[17] Lee Martin. 2015. The Promise of the Maker Movement for Education. *Journal of Pre-College Engineering Education Research (J-PEER)* 5, 1 (2015).

[18] Sylvia Libow Martinez and Gary Stager. 2013. *Invent to learn: Making, tinkering, and engineering in the classroom.* Constructing modern knowledge press Torrance, CA.

[19] Robert J Mislevy, Linda S Steinberg, and Russell G Almond. 2003. Focus Article: On the Structure of Educational Assessments. *Measurement: Interdisciplinary Research and Perspectives* 1, 1 (2003), 3–62.

[20] NGSS Lead States. 2013. *Next Generation Science Standards: For States, by States.* Number November. 1–103 pages.

[21] Seymour Papert. 1991. Situating Constructionism. In *Constructionism*, Seymour Papert and Idit Harel (Eds.). Ablex Publishing Corporation, Norwood, NJ, Chapter 1.

[22] Kylie Peppler, Erica Halverson, and Yasmin B Kafai. 2016. *Makeology: Makerspaces as learning environments.* Vol. 1. Routledge.

[23] Erica Rosenfeld-Halverson and Kimberly Sheridan. 2014. The Maker Movement in Education. *Harvard Educational Review* 84, 4 (2014), 495–504.

[24] Gery W Ryan and H Russell Bernard. 2003. Techniques to Identify Themes. *Field Methods* 15, 1 (2003), 85–109.

[25] Thomas A Schmitt. 2011. Current Methodological Considerations in Exploratory and Confirmatory Factor Analysis. *Journal of Psychoeducational Assessment* 29, 4 (2011), 304–321.

[26] Charles Teddlie and Abbas Tashakkori. 2006. A general typology of research designs featuring mixed methods. *Research in the Schools* 13, 1 (2006), 12–28.

[27] Shirin Vossoughi and Bronwyn Bevan. 2014. Making and tinkering: A review of the literature. *National Research Council Committee on Out of School Time STEM* (2014), 1–55.

[28] Peter S. Wardrip and Lisa Brahms. 2015. Learning Practices of Making: Developing a Framework for Design. In *Proceedings of the 14th International Conference on Interaction Design and Children (IDC '15).* ACM, New York, NY, USA, 375–378.

[29] Peter S. Wardrip and Lisa Brahms. 2016. Taking making to school. *Makeology: Makerspaces as learning environments* 1 (2016), 97–106.

[30] Mark Wilson. 2004. *Towards coherence between classroom assessment and accountability.* University of Chicago Press.