The Role of Evidence Centered Design and Participatory Design in a Playful Assessment for Computational Thinking About Data

Satabdi Basu[†]
SRI International
Menlo Park CA USA
satabdi.basu@sri.com

Betsy Disalvo
Georgia Tech
Atlanta GA USA
bdisalvo@cc.gatech.edu

Daisy Rutstein
SRI International
Menlo Park CA USA
daisy.rutstein@sri.com

Yuning Xu
Kidaptive Inc.
Redwood City CA USA
yuning.xu@kidaptive.com

Jeremy Roschelle
Digital Promise
San Mateo CA USA
jroschelle@digitalpromise.org

Nathan Holbert
Columbia University
New York NY USA
holbert@tc.columbia.edu

Abstract

The K-12 CS Framework provides guidance on what concepts and practices students are expected to know and demonstrate within different grade bands. For these guidelines to be useful in CS education, a critical next step is to translate the guidelines to explicit learning targets and design aligned instructional tools and assessments. Our research and development goal in this paper is to design a playful, curriculum-neutral assessment aligned with the 'Data and Analysis' concept (grades 6-8) from the CS framework. Using Evidence Centered Design and Participatory Design, we present a set of assessment guidelines for assessing data and analysis, as well as a set of design considerations for integrating data and analysis across middle school curricula in CS and non-CS contexts. We outline these contributions, describe how they were applied to the development of a game-based formative assessment for data and analysis, and present preliminary findings on student understanding and challenges inferred from student gameplay.

CCS CONCEPTS

• Social and professional topics~Computational thinking • Social and professional topics~Student assessment • Social and professional topics~K-12 education

KEYWORDS

Data and Analysis, CS assessment, Evidence centered design, Participatory design, Game-based assessment

[†]Corresponding author

ACM ISBN 978-1-4503-6793-6/20/03...$15.00
https://doi.org/10.1145/3328778.3366881

ACM Reference format:

Satabdi Basu, Betsy Disalvo, Daisy Rutstein, Yuning Xu, Jeremy Roschelle, and Nathan Holbert. 2020. The Role of Evidence Centered Design and Participatory Design in a Playful Assessment for Computational Thinking About Data. In Proceedings of the 2020 ACM SIGCSE Technical Symposium on Computer Science Education (SIGCSE'20), March 11-14, Portland, OR, USA. ACM, New York, NY, USA, XX pages. https://doi.org/10.1145/3328778.3366881

1 Introduction

Computing is an integral part of our world today and the public demand for computer science (CS) education is at an all-time high with schools continually seeking guidance on developmentally appropriate instructional material and assessments for CS. In particular, with non-CS teachers being asked to integrate a wide range of CS concepts in other disciplines, there are novel problems to address.

When the K-12 CS Framework was released in 2016 [18], it filled a definitional role in the field by providing guidance on what concepts students are expected to know and what practices students are expected to be able to demonstrate within certain grade bands. However, the framework does not define learning goals or describe examples of framework-aligned curricula and assessments. Hence, translating the guidelines in the framework to explicit learning and assessment targets and designing instructional material and/or assessments that align with the targets continue to remain critically important tasks in the field of CS Education.

We know of no assessment that directly targets the 'Data and Analysis' concept and all its sub-concepts (data collection, storage, visualization and transformation, and inference and models), as outlined in the K-12 CS framework. Some assessments like CT-STEM [23] and previous work in Math education [25] have assessed specific aspects of this construct. In this paper, we describe a principled approach for designing assessments for all the 'Data and Analysis' concepts for middle school students. We focus on elaborating learning (and assessment) goals for the data and analysis strand of the K-12 CS Framework, and designing learning experiences to support those learning goals, especially in non-CS classrooms.

With the amount of digital data in the world rapidly expanding, data and analysis constitute an essential ability required in all sectors of the global knowledge-based economy. Data literacy, encompassing collecting, managing, interpreting, evaluating and applying data in support of evidence-based decision making, is considered an important skill in all disciplines, not just CS [12]. Since data literacy is an important concept in a number of disciplines, we aimed to understand how both CS and non-CS teachers can integrate this concept into their courses, and how an assessment can provide useful feedback to teachers in various contexts. Using a design research approach, we worked with middle school teachers to identify a set of design considerations for incorporating data and analysis concepts into their courses and designing formative assessment tools for the same. Using the principled approach for designing data and analysis assessments and the design considerations we identified, we designed a game-based formative assessment for data and analysis in middle school. We report on data from a pilot study using the game to investigate middle school students' understanding and challenges with data related concepts. Our research is informed by the following research questions:

1. What should the assessment goals and assessment guidelines be in order to elicit evidence of competency with data and analysis, as outlined by the K-12 CS framework?
2. What are the design considerations when developing data and analysis activities and assessments for middle school CS and non-CS teachers?
3. How can the assessment goals and guidelines and design considerations be incorporated into a game-based formative assessment for Data and Analysis?

2 Design frameworks

To develop assessment goals for the CS concept of data and analysis, we used Evidence-Centered Design (ECD) [14], a principled assessment design framework that provides support to the validity argument that the assessment is measuring the intended constructs [13]. ECD promotes coherence in the design of assessment tasks and rubrics and the interpretation of students' performances by explicitly linking claims about student learning, evidence from student work products, and design features of tasks that elicit the desired evidence. ECD begins with a domain analysis, which in this case entails gathering and organizing information about the 'Data and Analysis' concept for grades 6-8. This is followed by domain modeling, which entails the articulation of a design specification for each construct of interest, which in turn informs the development of tasks and rubrics. A design specification is used to specify the focal knowledge, skills and abilities (FKSAs) or assessment targets corresponding to each construct, the type of evidence needed from the student to measure these FKSAs, and the characteristics and variable features of tasks that can be designed to elicit the evidence [15]. ECD has been used widely in CS, including development of the Principled Assessments for Computational Thinking [3, 7].

Determining what can and should be assessed also requires considering the learning context. To this end, we used participatory design (PD) [5, 6] - an established method in human computer interaction and an emerging method in the learning sciences [6] that allows researchers to involve stakeholders in the design of learning tools. In PD sessions, stakeholders start designing with scaffolded activities, that help them reflect upon design needs, assets and challenges, and provide inspiration for novel approaches to design [19]. Our use of PD follows in a tradition of using the design process as a research tool, a way to understand stakeholders better and to help them reflect and articulate issues that might impact design [24]. Our goal was to find ways to integrate data and analysis across the middle school curriculum, and PD helped create neutral spaces where various teachers' perspectives were considered in the design, regardless of their CS expertise. The use of design activities, where each teacher built from their own expertise, helped develop shared vocabulary among teachers and researchers, which in turn allowed them to speak freely and with greater understanding.

3 Methods

3.1 Developing assessment guidelines

We used the ECD framework to identify four design specifications, one for each sub-concept of the 'Data and Analysis' CS concept - Collection, Storage, Visualization and Transformation, and Inference and Modeling. Each design specification comprised target FKSAs to be assessed, potential observables including common errors to be looked for in student responses, characteristic or necessary features of assessment tasks, variable task features or ways a task could be varied, and additional knowledge, skills and abilities representing aspects that might be needed to complete tasks, but are not assessment targets.

For developing the FKSAs, we used the framework guidelines and current literature and also talked to experts in the field to elaborate on what it meant for students to engage with each of the data and analysis concepts in middle school. This included considering what knowledge and abilities students would have developed in previous grades, and what knowledge and abilities might be too advanced. Data Collection & Storage (**DCS**) FKSAs and Data Visualization, Transformation, & Inference (**DVTI**) FKSAs were identified as follows:

DCS1. Ability to identify variables or types of data that should be collected based on the purpose of the data collection.

DCS2. Ability to identify how to automate data collection (e.g. how often to collect the data, and the use of a computational tool to collect the data).

DCS3. Ability to identify an appropriate representation for the data that is to be collected and stored given the purpose of the data and the storage constraints (includes identifying types of metadata that might be collected).

DCS4. Ability to manage the tradeoffs between data collection and storage requirements.

DVTI1. Ability to create a visualization for a dataset.

DVTI2. Ability to identify which data should be used to address a certain question. This includes identifying outliers and creating rules for the computer to filter outliers.

DVTI3. Ability to transform data to highlight a specific relationship.

DVTI4. Ability to use the data to create an appropriate model that demonstrates relationships within the data.

DVTI5. Ability to interpret data models and visualizations for making predictions or drawing conclusions.

DVTI6. Ability to refine a data model using new or additional data (includes the knowledge to go back to the model to see if it still fits with new data).

Once the FKSAs were developed, they were reviewed for content coverage and grade-level appropriateness and a subset was then selected as the focus of our game-based assessment – **DCS1**, **DCS2**, **DCS4**, **DVTI2**, **DVTI5**. For each FKSA, further information was specified regarding evidence needed to measure the FKSA and what features tasks should have to ensure this evidence can be gathered. This helped articulate a set of Assessment Guidelines (**AG**) for designing assessments that elicit evidence of the FKSAs listed above. AG1-AG4 aligns with the DCS FKSAs while AG5-AG8 align with the DVTI FKSAs.

AG1. Students must be given the purpose behind collecting and/or storing data.

AG2. Students must be given opportunities to choose aspects of data collection (e.g. what variables, frequency of data collection, the format of storing the variables).

AG3. Vary the number of and type of choices students have for data collection and storage to vary the task complexity.

AG4. Scores should be based on the appropriateness of student choices based on the purpose of the data collection.

AG5. Students should be provided with a dataset or data visualization(s) and a purpose or question to address.

AG6. Students must be given opportunities to develop a model or use a model.

AG7. Vary the complexity of the data, the visualization used, and the data model relationships to vary task complexity.

AG8. Scores should be related to the appropriateness of the data representation used or generated and/or the appropriateness of the inference made based on the data.

These FKSAs and AGs help answer our 1st research question and are an important step in the design of our playful formative assessment. We believe that they can provide valuable insights to K-12 CS educators and researchers and can be reused for developing additional assessments or for aligning instructional material to the guidelines for the 'Data and Analysis' strand. Our design process for generating FKSAs and corresponding assessments also has implications for K-12 CS curriculum and assessment design since it can be generalized for other concepts and grade bands in the CS framework.

3.2 Developing Design Considerations

In parallel, we conducted PD research to develop design considerations guided by insights from middle school teachers. The research took place in three sessions with 11 teachers who represented a range of content areas (Literature, Math, CS, Science, Art, Music and Health), and teaching experience (1 year to 39 years). Teachers were guided through design activities relevant to integrating and assessing data and analysis concepts in their classrooms.

In the first activity, teachers were given a set of small colored post-it notes representing the four data related sub-concepts in the K-12 CS Framework (Figure 1). Teachers were asked to use these post-it notes along with a template to design a data and analysis lesson in a current course they teach. After they completed the lesson design, they were given a set of stickers with common assessment methods (e.g. summarization, quizzes, reflection, peer feedback) and blank stickers to write their own assessment methods, and asked to design a way to conduct formative assessment of students' data and analysis learning. Teachers then walked through the lesson and formative assessment designs, sharing with the group.

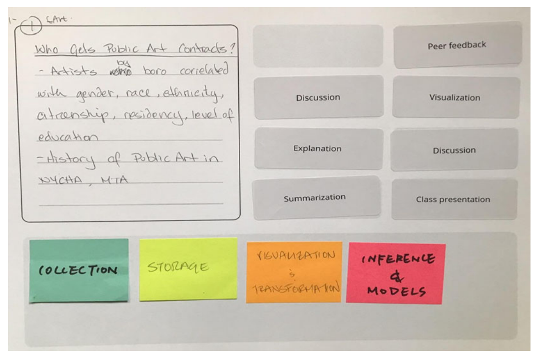

Figure 1. An art teacher's activity and assessment design for teaching data and analysis

All the teachers, with the exception of the CS teacher, revealed that they had hesitations about teaching CS concepts more generally, but found data and analysis concepts more accessible than other concepts outlined in the CS Framework. In the first activity, most teachers recognized current classroom activities, or quickly brainstormed new activities, that taught data and analysis and still tied into the subject matter they were teaching. The teachers who expressed the most enthusiasm for integrating computing were also those with concerns that the heavy emphasis on STEM will result in reduced funding for their classes (Art, Health, Music). The health teacher had already worked with his school's CS teacher to develop a semester long activity focused on health data to promote data literacy. On the other hand, while we initially identified more overlaps of data related learning goals with Math and Science, teachers of these subjects were less able to imagine integrating computing due to the many learning goals they currently need to address. The least enthusiastic to incorporate CS were the History and English teachers, who told us they felt over-burdened, and ensuring kids can read and write were higher priorities than teaching data and analysis.

All teachers agreed that assessment was a challenge. Teachers in Art, Health, and Music self-identified as needing help to get a data and analysis project started and to evaluate if students are meeting learning goals. As one Art teacher told us, "There are so many ways that student won't get it but I'm not really sure what those ways might be." Interestingly, teachers chose stickers for time intensive assessment techniques such as presentations, written reflections, and visuals. They seemed to avoid less time intensive quizzes or class polling techniques in an attempt to reduce the testing burden and possible adverse impacts on more high stakes testing of other subjects.

In a second activity, teachers were shown a video walk-through of a game, *Plague Inc.* [20], that uses a number of different data representations and data sources, and were asked to think about whether players learned about data and analysis. After viewing the game, teachers were asked to provide feedback on the possible use of games in their classrooms and what game-data they would find useful for assessment. We allowed them to play with the *Plague Inc.* game to help them identify what sorts of behavior might indicate learning of data related skills. We then asked them how they might like that data presented to them on a dashboard.

In this activity, teachers were unsure about measuring learning when watching the walk-through. When we asked where teachers might implement a similar playful assessment, they suggested designing short experiences, 5 – 20 min, that they could use at the end of a class period or as a reward for the class. As for feedback about student behavior, teachers requested both real-time feedback to ensure students were engaged during class and feedback at the end of the day or week providing more detailed information about student learning. Teachers also cautioned against formative assessment data being used as performance evaluation measures, much like standardized tests. Finally, teachers expressed concerns that games like *Plague Inc.* would not keep student data secure, would have issues running on the lower bandwidth available in many schools, and would not work on the devices (Chromebooks) available in their classrooms.

Overall, the PD activities helped answer our 2nd research question and resulted in four design considerations (DC):

DC1. Integration of data and analysis across various curricula is a rich opportunity, particularly with subjects such as Art, Music and Health.

DC2. Assessments of CS concepts are needed; however, they should be short and playful and not cause testing fatigue that traditional methods might.

DC3. The form of assessment feedback should include just-in-time behavior data, in addition to aggregated classroom learning outcomes that explain what the findings mean for CS learning and how to use the feedback. In addition, assessment feedback should not be merely numerical to avoid being used as a teaching performance measure.

DC4. Digital tools should be built keeping in mind available technology, technology limitations, issues of student privacy and network bandwidth.

3.3 Game design and piloting

We used the FKSAs and assessment guidelines derived using an ECD approach and the design considerations derived using a PD approach to then design *Beats Empire*, a game-based formative assessment for the middle school 'Data and Analysis' strand of the K-12 CS framework. Games can provide students with opportunities to encounter and use concepts and practices in authentic and meaningful settings [10, 21]. They have also been shown to be helpful for assessing and understanding CS understanding in particular [1, 22]. Further, the game format supports our findings that short and playful assessments are necessary when integrating CS across the curriculum (**DC2**). *Beats Empire* has been designed as a web-based stand-alone assessment tool that is not subject matter dependent and is a light-weight addition to regular classroom activities. It can be used in multiple short sessions or fewer longer sessions so that teachers can adapt it to their needs.

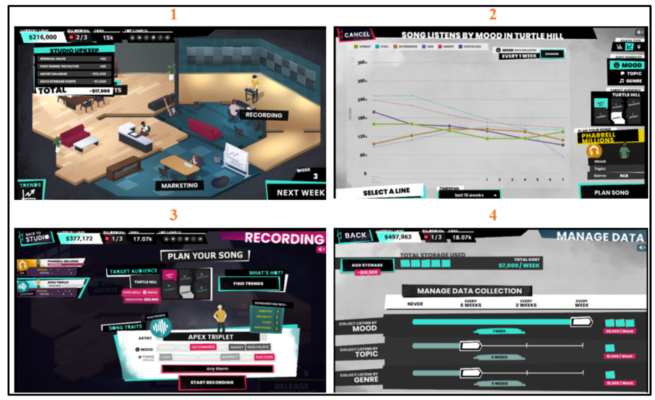

Figure 2: Screenshots from 'Beats Empire' – 1: studio home screen, 2: song popularity trends screen, 3: song recording screen, 4: data collection and storage screen

Beats Empire is a music studio management game where players collect data about listener interests and use this data to make decisions about which artists to sign and what songs to record (Figure 2). Players are free to define their own goals as they play the game (e.g. releasing the most #1 songs, creating a studio that specializes in a particular song genre, etc.). However, in every case, players must thoughtfully decide what data to collect, how to interpret the data, and what decisions to make based on the data. "Management" genre games are known to engage players in highly mathematical practices that include reading, analyzing, and acting on large amounts of data and advanced representations to manage complex interconnected systems, and are hence highly compatible with our assessment needs. By using music as the context and leaving the game play goals more open, the game can be integrated across a number of different subject areas and resonate with teachers that are particularly interested in integrating CS concepts, such as Art and Music teachers (**DC1**).

Within the context of the game, it was important that students be given the opportunity to engage with the FKSAs we identified. We knew we wanted to measure students' ability to both identify which data should be used to address a certain question (**DVTI2**) and to interpret visualizations to make a prediction or draw a conclusion (**DVTI5**). This provided us with the guidance that we would need to allow students to have a choice regarding which data representation they use (and whether or not they use data at all) as well as an incentive for correctly interpreting the data representation. When exploring data representations for the game, the fact that there were teachers outside of CS who might use the game (**DC1**) led to the exclusion of certain data representations such as databases that would require queries. To address these FKSAs, we applied **AG5** and decided to present students with multiple data representations of song popularity (line graphs, bar graphs, heat maps) and focus on students' selection of appropriate representations for making decisions as opposed to having students create representations. Students are able to use the data shown in the graphs/representations to make decisions about hiring artists and recording songs, as well as make predictions about the data in the game. Students are rewarded in the game for choosing a song that is high in popularity, and correct predictions result in addition money and followers.

As for DCS, we decided to focus on measuring students' ability to identify the types of data to collect (**DCS1**) and how often to collect the data (**DCS2**) and ability to manage tradeoffs between data collection and storage requirements (**DCS4**). We de-emphasized DCS3 since DCS3 deals with various storage formats like digital, analog, ASCII, and UNICODE that might not be suitable for non-CS teachers. Applying **AG1** and **AG2**, we designed the game to initially provide students with a limited amount of song data, and later provide opportunities to collect/purchase additional data. Students will need to decide what additional data they want to collect to move forward, and how often they want to collect the data. Students are given a limited amount of money and storage to encourage them to make thoughtful decisions about the data they are collecting and how that relates to the amount of storage they have.

We have recently begun developing a teacher dashboard that provides just-in-time data on student engagement as well as information on what the game data means in terms of learning outcomes. Based upon the PD findings (**DC3**), the data is expressed in heat maps, rather than numerical scores, to avoid the impulse of teachers to use the outcomes as performance measures. In addition, *Beats Empire* is designed to consider teachers' technical and privacy concerns (**DC4**) and runs seamlessly on inexpensive laptops such as Chromebooks, keep student data secure, and work with intermittent internet.

Pilot study using Beats Empire

We conducted a preliminary study using *Beats Empire* with a group of 28 middle school students (gender-balanced) in a large, diverse city in the U.S. Students engaged in 45 minutes of in-class gameplay over 2 days and participated in a post-game focus-group discussion. We investigated students' gameplay logs and focus group discussions to elicit evidence for the target FKSAs. Our primary goal was to see if the game would be meaningful and engaging to students, and if it could produce evidence for or against our assessment goals or FKSAs.

4 Results from pilot study using Beats Empire

Overall, all students enjoyed playing the game and were highly engaged. Several students played multiple sessions of the game, even after winning previous sessions, and many played the game at home in addition to the time played in class.

One important finding was that students exhibited different degrees and types of data usage. Several students used data about artists' talent and reliability as well as data about popularity of song characteristics to make their in-game decisions. But, some students used their understanding of popular culture (making connection between artists' game-names and real-life artists) as well as their preferred music genres to make game-based decisions. We found that students recorded 1425 songs in all, with most students recording between 20 and 50 songs. 43% songs were recorded without looking at any data (bar or line graphs) about the popularity of song characteristics; 11% songs were recorded after accessing data, but the choices were not data-supported; while 46% choices were data-supported (**DVTI5**). 44% song choices were supported using the bar graph alone, less than 1% choices supported using the line graph alone, and 1% of song choices were supported using both bar and line graphs. Interpreting line graphs was challenging for most students, and during the focus group, one-third students suggested that they used line graphs to determine which song characteristic was most popular rather than to identify trends over time (**DVTI2**).

We also found that students were not consistent about their use of data over time. Broadly, we found that some students rarely used data, some students increased their use of data as they played the game, and some students consistently used data. This also led to discussions about how to provide feedback to teachers about the data. In keeping with **DC3**, we decided to provide feedback about the classification of students based on their use of data as opposed to details about exactly how many songs students recorded that were a result of meaningful data-based decisions.

For data collection and storage, most students deferred to the game defaults and chose not to explicitly change the variables on which data was being collected or the frequency at which data was being collected (**DCS1**, **DCS2**). Only 5 of 28 students chose to vary the frequency of data collection and they all demonstrated an awareness of the relations between data collected and storage required (**DCS4**) – for example, buying storage along with increasing collection frequency.

Overall, our results show that our game was engaging to students while being able to elicit evidence of students' understanding of our target FKSAs. About half the students were consistently able to draw meaningful inferences from data visualizations to decide what type of songs to record in which location (**DVTI5**), and a few students showed improvement in this FKSA over time. Most students were not proficient with the use of line graphs and had difficulty distinguishing the utility of line and bar graphs (**DVTI2**). Few students demonstrated an awareness of the relations between data collected and storage required (**DCS4**).

5 Discussion and implications for pedagogy

This paper describes the convergence of an ECD approach for assessment development and a PD approach for identifying teacher needs and viewpoints to develop a game-based formative assessment for middle school students engaging with the Data and Analysis strand of the CS framework.

The first primary contribution of this work is the creation of assessment goals and guidelines that provide teachers and researchers with concrete learning and assessment goals for middle school data and analysis. We describe a principled approach for transforming the guidelines in the K-12 CS framework for the 'Data and Analysis' concept to specific learning (and assessment) targets. The FKSAs can be used by K-12 educators and researchers for designing instructional material aligned to guidelines in the CS framework and other forms of data-related assessments, for example, traditional paper-pencil based summative assessments, or assessments requiring data manipulation in a science context. Finally, our principled ECD process has broader implications for K-12 CS curriculum and assessment design beyond the realm of the 'Data and Analysis' strand since it can be generalized for other CS concepts and grade bands.

Second, our PD sessions with middle school teachers (CS and non-CS) provide valuable information about subject areas (Art, Health and Music) where new CS content around data might be best received, and teachers' viewpoints and priorities when considering integrating data and analysis activities and assessments in their classes. These insights helped guide our assessment design and formative feedback mechanisms. Most design considerations identified in this paper have transferable implications for CS integration across curricula.

Third, we apply our findings from the ECD and PD processes to develop a curriculum-neutral game-based formative assessment for middle school data and analysis. We argue for more authentic assessments that focus not just on students' knowledge of data related concepts but also on students' ability to engage with and apply the concepts in authentic contexts. Our initial results indicate that several students had difficulty distinguishing the utility of line and bar graphs, and relating data collected with storage required, pointing to the need for emphasizing these concepts in curricular initiatives.

Engaging students in real-world relatable experiences makes *Beats Empire* an assessment that fits into classrooms across subject areas and is less likely to cause testing fatigue. However, it also makes the data harder to reliably interpret and generate actionable feedback for teachers. For example, while some gameplay actions are highly indicative of proficiency on data-related measures, does lack of such actions necessarily indicate lack of proficiency? Based on both log data and focus groups, we conclude that lack of indicative gameplay actions in *Beats Empire* does not necessarily indicate lack of proficiency on data-related measures, and information on students' gameplay can point teachers to follow-up conversations that may help students reflect on their in-game actions. The culturally relevant context of music means that students often incorporate understandings derived from their personal lives related to the music industry into their decision-making process in the game. For example, one student explained that the in-game artists had similar names to real-world artists, and she decided what type of song to record based on what songs the artists generally make. Another student justified recording songs about 'Love' because she preferred such songs. Such students require further probing on their data-based decision-making abilities through follow-up activities. Additionally, some students use data and later drop off due to various reasons like game exploration, changing goals, etc. We would still consider them as data users due to their initial evidence.

We are currently working on next steps that include piloting the game in more classes with more students and teachers, designing a teacher dashboard to support teachers in implementing *Beats Empire* in the classroom, and developing supporting teacher materials and follow-up activities for a diverse cohort of middle school teachers. The goal of formative assessments is not assigning a grade to each student or assessing teaching skills, but rather providing actionable feedback to teachers to help adapt their instruction. To that end, *Beats Empire* can act as a shared object to think with, helping contextualize classroom discussion around data. We are also developing follow-up classroom activities and teacher guides to help teachers use information from the game to continue conversations around data literacy with their students and probe further into student abilities and challenges. While there might not always be enough information from the gameplay alone to make strong conclusions about student abilities, the information can guide teachers on how to adapt instruction and effectively use the suggested follow-up activities to initiate meaningful class discussions. A formative assessment tool can realize its full potential only when we support teachers in interpreting and responding to the formative feedback.

ACKNOWLEDGMENTS

We would like to acknowledge the National Science Foundation (NSF Award #1741956 and Award #1742011) for funding this research. We are grateful to the middle school teachers who shared their insights with us, the students who participated in our pilot studies, and to all the project team members who helped conduct the interviews and pilots.

REFERENCES

[1] Berland, M. (2016). Making, tinkering, and computational literacy. In K. Peppler, E. Halverson, & Y. B. Kafai (Eds.), *Makeology: Makers as Learners* (Vol. 2, pp. 196–205). NYC: Routledge.

[2] Berland, M., Martin, T., Benton, T., Petrick Smith, C., & Davis, D. (2013). Using Learning Analytics to Understand the Learning Pathways of Novice Programmers. *Journal of the Learning Sciences*, *22*(4), 564–599. https://doi.org/10.1080/10508406.2013.836655

[3] Bienkowski, M., Snow, E., Rutstein, D. W., & Grover, S. (2015). Assessment design patterns for computational thinking practices in secondary computer science: A First Look. Menlo Park, CA: SRI International.

[4] Cuban, L. (1993). How teachers taught: Constancy and change in American classrooms, 1890-1990. Teachers College Press.

[5] DiSalvo, B., & DiSalvo, C. (2014). Designing for democracy in education: Participatory design and the learning sciences. In *Proceedings of the Eleventh International Conference of the Learning Sciences (ICLS 2014)*. Boulder, CO: International Society of the Learning Sciences.

[6] DiSalvo, B., Yip, J., Bonsignore, E., & Carl, D. (2017). Participatory design for learning. In *Participatory Design for Learning* (pp. 3-6). Routledge.

[7] Goode, J., Chapman, G., & Margolis, J. (2012). Beyond curriculum: the exploring computer science program. *ACM Inroads*, *3*(2), 47-53.

[8] Greenberg, Saul. (2002, University of Calgary CPSC 481 coursenotes edition) "Working through Task-Centered System Design." In Diaper D. and Stanton, N. (Eds) *The Lori Shyba • 20 Handbook of Task Analysis for Human-Computer Interactions*. Lawrence Erlbaum Associates.

[9] Holbert, N., & Wilensky, U. (2018). Designing Educational Video Games to Be Objects-to-Think-With. *Journal of the Learning Sciences*, *0*(ja), null. https://doi.org/10.1080/10508406.2018.1487302

[10] Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (pp. 4–4).

[11] Laughey, D. (2006). *Music and Youth Culture*. Edinburgh University Press.

[12] Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, *42*(1), 30-37.

[13] Mislevy, R.J. (2007). Validity by design. *Educational Researcher, 36*(8), 463-469.

[14] Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6–20.

[15] Mislevy, R. J., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of test development* (pp. 61-90). Mahwah, NJ: Lawrence Erlbaum.

[16] Papert, S., & Harel, I. (1991). Situating constructionism. In S. Papert & I. Harel (Eds.), *Constructionism* (pp. 1–11). New York: Ablex Publishing.

[17] Papert, S. (1980). Mindstorms: Children, computers, and powerful ideas. Basic Books, Inc.

[18] Parker, M. C., & DeLyser, L. A. (2017). Concepts and Practices: Designing and Developing A Modern K-12 CS Framework. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (pp. 453-458). New York, NY, USSA: ACM.

[19] Sanders, E.B.N., and Stappers, P.J. Convivial design toolbox. 2012. Publishers Group UK, Amsterdam.

[20] Public Health Matters Blog: Plague Inc., 2013, accessed May 29, 2014. [Online]. Available: http://blogs.cdc.gov/publichealthmatters/2013/04/plague-inc/

[21] Shaffer, D. W., & Gee, J. P. (2012). The right kind of gate. Technology-Based Assessments for 21st Century Skills: Theoretical and Practical Implications from Modern Research, 211–228.

[22] Weintrop, D., Holbert, N., Horn, M. S., & Wilensky, U. (2016). Computational Thinking in Constructionist Video Games: *International Journal of Game-Based Learning*, *6*(1), 1–17. https://doi.org/10.4018/IJGBL.2016010101

[23] Weintrop, D., Beheshti, E., Horn, M. S., Orton, K., Trouille, L., Jona, K., & Wilensky, U. (2014, July). Interactive assessment tools for computational thinking in High School STEM classrooms. In International Conference on Intelligent Technologies for Interactive Entertainment (pp. 22-25). Springer, Cham.

[24] Zimmerman, J., Forlizzi, J., & Evenson, S. (2007, April). Research through design as a method for interaction design research in HCI. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 493-502). ACM.

[25] Lehrer, R., Kim, M., Ayers, E., & Wilson, M. (2014). Toward Establishing a Learning Progression to Support the Development of Statistical Reasoning." In A. Maloney, J. Confrey, & K. Nguyen (Eds), Learning over Time: Learning Trajectories in Mathematics Education. Charlotte, N.C.: Information Age Publishers.