Game-Assisted Assessment for Broader Adoption: Participatory Design and Game-Based Scaffolding

Evan Rushton and Seth Corrigan
SNHU Innovation Center, San Francisco, CA, USA
evan.rushton@gmail.com
seth.corrigan@snhu.edu

Abstract: 21st Century Standards and the Deeper Learning movement emphasize the ability to think critically and solve complex problems, to work well in teams, and to communicate effectively. While traditional classroom activities can meet these objectives, digital games and simulations provide unique affordances. When designed to incorporate formative assessment functions, games and simulations can capture detailed data on learners' performances and provide learners with immediate feedback. In spite of their strengths, barriers exist to practitioners' adoption of game-based and simulation-based formative assessments. Adoption can be slowed where product designs do not account for unique local requirements of classrooms and schools. The current work investigates reduction and removal of barriers to adoption of games and simulations among classroom instructors through use of the Integrated BEAR Design System (IBDS). The IBDS provides a design process that accounts for local requirements by engaging practitioners in principled design and development of game-based formative assessments. The paper summarizes the IBDS and a single case in which the IBDS was applied to design a game-based formative assessment for collaborative-problem solving, Little Fish Lagoon. The game is accompanied by a stand-alone chat system, Libra Text, that allows collaborating players to send text messages to each other while they use the game. Study participants were six instructors from six U.S. schools. The participating instructors planned for broad adoption of the multiplayer collaboration game in their local classroom settings. The authors illustrate their use of the IBDS with the participating instructors in order to co-develop formative assessments that fit their local needs using data collected from the Little Fish Lagoon educational game and the Libra Text chat tool. The benefits of the IBDS, its implications for learning designers, potential improvements, and needed future research are discussed. The paper is expected to be of interest to learning and assessment designers working with educational games and simulations, and others interested in barriers to adoption of new technologies in general.

Keywords: Formative Assessment, Collaboration, Participatory Design, BEAR Assessment System, Barriers to Adoption, Game-Based Learning

1. Introduction

21st Century Skills describe a broad range of knowledge, skills, attitudes and beliefs - such as problem solving, critical thinking, communication, collaboration, and self-management. Consistent with similar statements from representatives of the OECD, the National Research Council of the U.S. (2012) has recommended, "sustained support for the development of valid, reliable, and fair assessments of intrapersonal and interpersonal competencies." Interpersonal competencies include teamwork and collaboration.

New technologies such as digital games and simulations can provide rich contexts for learning experiences that require and support development of collaboration and problem-solving (Behrens, DiCerbo & Foltz, 2019; Graesser et al., 2018). The novelty of digital games and simulations for teaching and learning however can also serve to inhibit their adoption (Hamari & Nousiainen, 2015; Stieler-Hunt & Jones, 2017; Watson & Yang, 2016). This problem is made worse when the new technologies are designed in ways that do not account for local school and district contexts (Jean Justice & Ritzhaupt, 2015; Sánchez Mena & Martí Parreño, 2017).

This article presents the Integrated BEAR Design System (IBDS) – an approach to collaborative design that aims to reduce and remove barriers to adoption of new technologies such as educational games and simulations by making them more responsive to local needs. The IBDS incorporates existing work in Participatory Design (PD) and extends the BEAR Assessment System (BAS) (Wilson & Sloane, 2000) to create a flexible design approach that can incorporate requirements stemming from local contexts.

The study's teacher-participants utilized the Integrated BEAR Design System (IBDS) to co-develop formative assessments that fit their local settings using data collected from the Little Fish Lagoon educational game and the Libra Text chat tool. Little Fish Lagoon is a game-based formative assessment of collaborative problem-solving skills that puts players in charge of managing a North Atlantic Fishery. It is accompanied by the stand-alone texting system, Libra Text, which permits collaborating players to send text messages to each other as

Reference this paper: Rushton, E., and Corrigan, S., 2021. Game-Assisted Assessment for Broader Adoption: Participatory Design and Game-Based Scaffolding. *The Electronic Journal of e-Learning,* 19(2), pp. 71-87, available online at www.ejel.org

they use the game. The game and texting system, are modular in that Libra Text can be used either with or without the Little Fish Lagoon game. Instructors can use the stand-alone chat application in other lessons throughout the school year in order to facilitate online collaborations and monitor collaboration skills during group work.

2. Background

It is difficult to adopt new technologies when they do not meet requirements of local classroom and school contexts. As a means to improving adoption of new technologies for formative assessment and learning, the IBDS incorporates existing work in Participatory Design (PD) and extends the BEAR Assessment System (BAS) (Wilson, 2009) to create a flexible design approach that can incorporate requirements stemming from local contexts. In what follows, PD and the BAS are summarized in order to describe how the IBDS incorporates both and provides a process for accounting for local contexts.

2.1 Participatory Design

Participatory Design (PD) involves non-designers in co-design activities throughout the design process. The authors share Brandt's view that, "Designing the design process itself is just as important as designing the artefact (2006)." Brandt, Sanders, and Binder (2012) describe participatory tools and techniques as the "scaffolding for a temporary community of practice in the making." Their conceptualization of Participatory Design (PD) as a 'third space' that belongs neither to the participants nor designers is formalized in Drain and Sanders (2019) as the collaboration system model. This holistic view of the designer-participant collaboration considers the society and culture in which it occurs, the environmental influence caused by the activities planned and facilitated by the designer, the designer's knowledge and activities, and the participants' knowledge and capacity to participate.

The chosen PD framework has three dimensions: form, purpose and context (Sanders, Brandt & Binder, 2010). *Form* describes the kind of action that is taking place between participants including making, telling and enacting. *Purpose* describes why the tools and techniques are being used including probing, priming, understanding current experience, or for generating ideas about the future. *Context* describes where and how the tools and techniques are used including group size and composition, face-to-face vs. online, venue, and stakeholder relationships. Table 1 illustrates these three dimensions and their descriptors. In the Design section of this article the IBDS is described along these dimensions. The motivations for iterating this process are discussed in the Discussion section.

Table 1: Dimensions of the Participatory Design Framework

FORM	PURPOSE	CONTEXT
Making	Probing	Group size and composition
Telling	Priming	Face-to-face vs. online
Enacting	Understanding current experience	Venue
	Generating ideas about the future	Stakeholder relationships

2.2 BEAR72 Assessment System

The BEAR Assessment System (BAS) is a principled approach to assessment design that supports integration of assessment into the classroom teaching and learning process (Wilson & Sloane, 200072). It is grounded by four principles: (1) a developmental perspective of student learning; (2) a clear match between assessment and instruction; (3) management by teachers; and (4) sound standards of validity and reliability. Each principle is operationalized (Kennedy, 2005) by an associated building block of the design process. These building blocks include: (i) construct map, (ii) items design, (iii) outcome space, and (iv) a measurement model. At a high level, instruction should be informed by ongoing formative assessment, with use of assessment items that reflect well the teaching and learning goals within the classroom. The BAS principles and building blocks are displayed in Figure 1.

Figure 1: The building blocks and principles of the BEAR Assessment System (BEAR Center, 2014)

Wilson and Sloane describe the first principle of the BAS, a developmental perspective of student learning, as "an approach that focuses on the process of learning and on an individual's progress through that process." (2000) Therefore, a *construct map* defines a continuum of qualitatively different levels of sophistication for the knowledge, skill, or ability one wishes to measure. The second principle is to clearly match formative assessment and instruction. Formative assessment *items are designed* to target various levels along the continuum of the construct and student progress is monitored to inform which assessment items and tasks are appropriate throughout the course of instruction. Teachers use the evidence specified in the *outcome space* to assess student performance, set performance standards, track progress over time and provide meaningful feedback. The coherence provided by the outcome space allows for management by teachers, the third principle of the BAS. Once teachers, subject-matter experts, and designers have created a construct map with evidence criteria defining the outcome space and are designing formative tasks that target specific criteria, a *measurement model* can be chosen to uphold standards of validity and reliability, the fourth principle.

The Integrated BEAR Design System (IBDS) extends the BAS by engaging practitioners working in diverse teaching and learning environments in order to ensure the resulting design products meet requirements stemming from teachers' contexts. This article focuses on the co-design of the building blocks of the BAS with teachers across six contexts in the U.S. during the 2019-2021 school years.

2.3 The IBDS

Previous attempts at assisting practitioners in adopting immersive digital experiences have shown that guidance from experienced coaches, building community, and personal support contribute to practitioner uptake (Stieler-Hunt & Jones, 2019). Four main areas of competence have been identified in game-based pedagogy: pedagogical, technological, collaborative and creative (Nousiainen et al., 2018). Our participatory design process incorporates all of these elements and we hypothesize it will increase ease of use and usefulness by targeting the external factors of developing instructional support materials and broadening the set of subjects and topics covered by a single game; and the internal factor of increasing compatibility with teaching methods. The means to achieve these ends are co-development of materials with teachers across diverse settings and game-assisted formative assessment tailored to local needs.

3. Methods

3.1 Teacher recruitment

The teacher advisory was formed in Winter 2020 with six teachers located in three states across the U.S. The opportunity was shared during Fall 2019 with a flyer, written announcement, and interest form that were circulated through popular social media channels and emailed to personal contacts involved in game-based learning. Advisors were selected from a recruitment pool of 25 teachers. All of the teachers offered spots accepted the role. The selection was based on teachers with common discipline areas and contextual diversity. Two of the project's teacher-advisors teach high school mathematics, three teach high school biology, and one teaches middle school gaming and coding. They all come from different schools. Three schools are ranked in the top 30% for their states while the other three are ranked in the bottom 50% for theirs. Half of the schools have a majority of minority students and the number of students per grade level ranges from 120 to 880. These and additional descriptive data are provided in the Appendix. Teacher advisors signed agreement forms to a schedule of work with competitive compensation.

3.2 Game and communication visualization development

3.2.1 Little Fish Lagoon

A collaboration game entitled Little Fish Lagoon that collects data on patterns of communication amongst three players was developed during Fall 2019. Note the right pane of the interface shown in Figure 2 contains a chat window and communication visualization while the main lagoon window is the game area. The goal is to introduce communication tools and collaboration assessment components in a game context that requires collaborative activity to win.

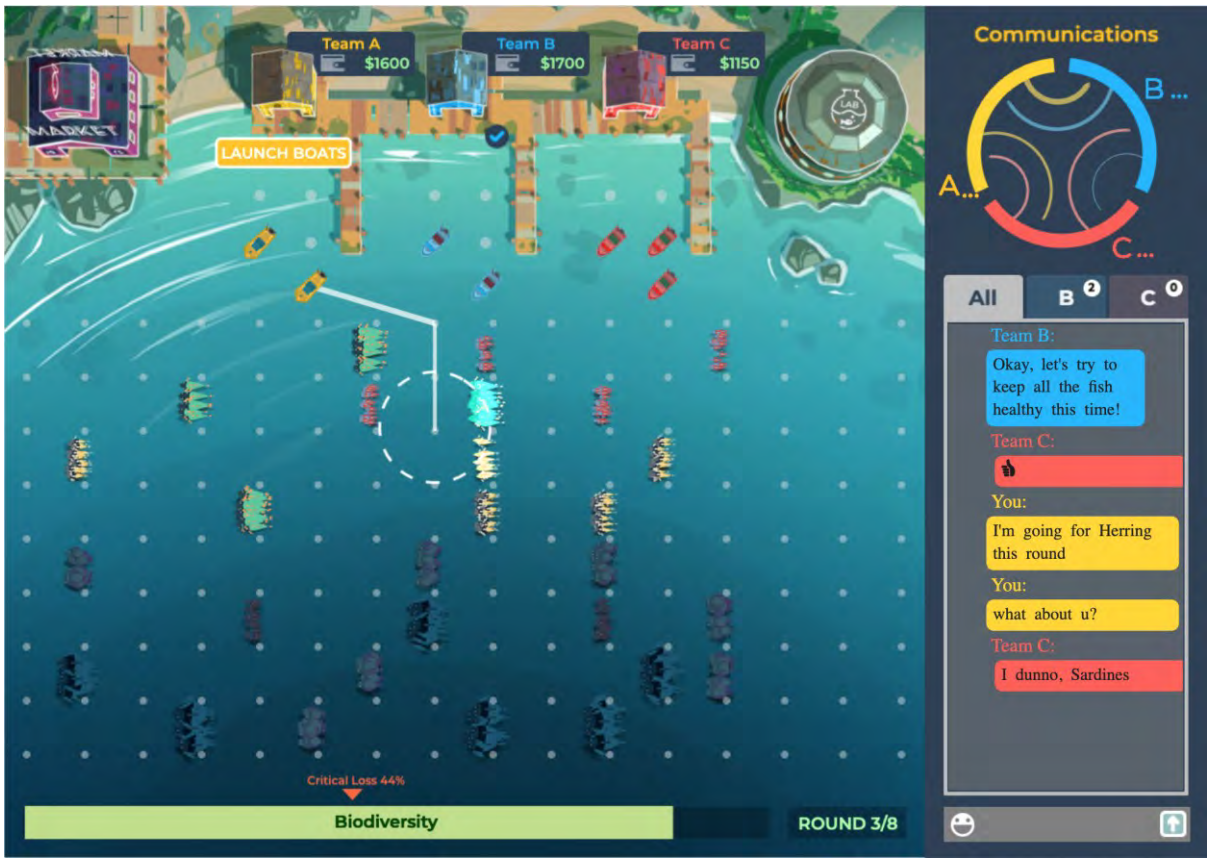


Figure 2: Player interface for Little Fish Lagoon

Players compete against each other within a common pool of resources they have to collectively manage. Some amount of cooperation and communication about the system they are working within is necessary to win the game. Each team directs their boats to catch one of the species of fish in the harbour in order to generate revenue to add to their score and upgrade their fleet. The challenge arises when certain fish populations begin to dip and the whole ecosystem becomes increasingly unstable. Players need to coordinate

their actions and weigh maximizing their own profits versus keeping the system sustainable. A player can either send a message to one or both of the other two players. The communication visualization in the top right corner of the player interface updates in real-time when players text each other. It provides a representation of who they are communicating with and how often.

3.2.2 Little Fish Lagoon teacher dashboard

The teacher dashboard was developed to gather and display communication and progress data collected by Little Fish Lagoon during Spring 2019. The class overview page in Figure 3 displays each group's visualization, current round, and biodiversity score. The goal was to surface data for a teacher to manage gameplay and collect evidence about individual and group collaboration skills.

The communication visualization displays the relative frequency of messages sent by each player. Note the varying thicknesses of the directed arrows between players A, B, and C in Figure 3. Patterns in the data emerge that help inform the player or teacher about the communication dynamic. In Figure 3 for example, group 1 shows players A and B communicating privately and not involving player C, group 2 shows a balanced communication pattern among all three players, and group 3 shows player C sending most of the messages.

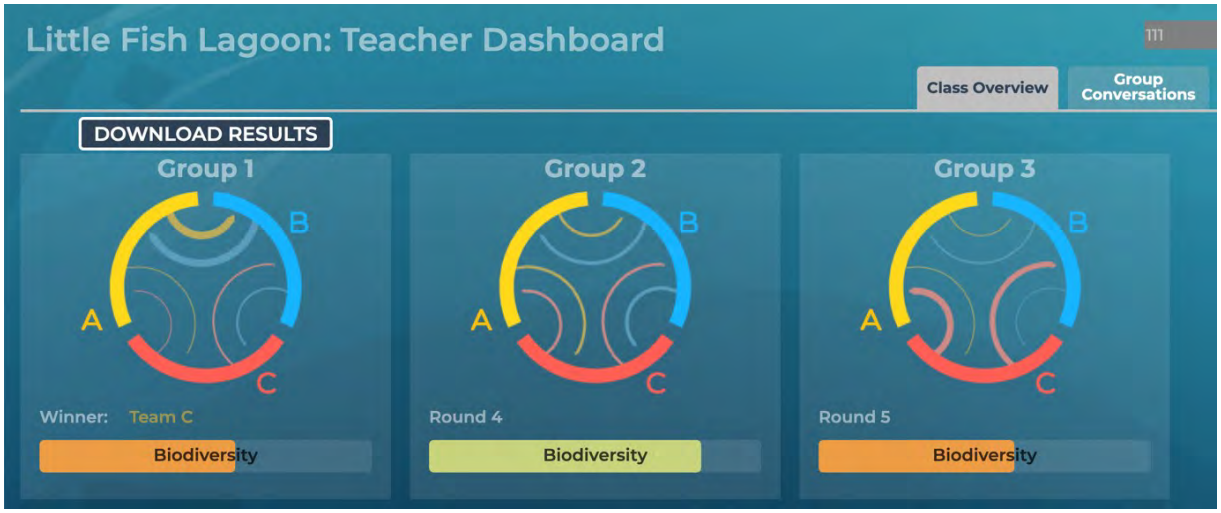


Figure 3: The class overview page for Little Fish Lagoon's teacher dashboard

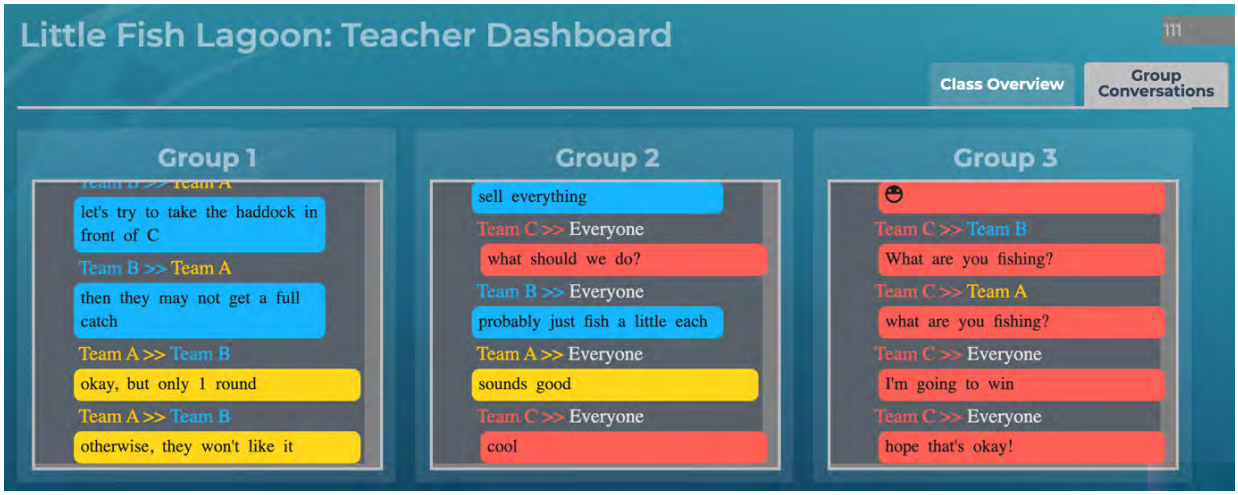


Figure 4: The group conversations page for Little Fish Lagoon's teacher dashboard

The group conversations page in Figure 4 displays the chat logs for each group. Messages from the chat logs can be used as evidence fragments to support or dispute claims made by self- and peer-ratings on criteria for collaboration and/or problem solving skills as described in the measurement model section. The goal in providing these data to teacher-advisors was to co-design formative assessment items and protocols that would support students to develop these 21st Century skills.

3.2.3 Libra and Libra teacher dashboard

The communication tool and visualization were extracted from the game and generalized for use with any classroom activity in a free standalone chat application for chrome entitled Libra during Fall 2020. It serves as a backchannel for collaborative activities done with groups of three. It collects and displays the same data as the in-game chat, but instead of tracking game progress it displays chat frequency in a stacked horizontal bar below the weighted directed graph as shown in Figure 5a. The Libra teacher dashboard illustrated in Figures 5b-c displays group data in the same manner as the Little Fish Lagoon teacher dashboard.

Figure 5: Interface for Libra (a), the Libra teacher dashboard's room overview (b) and group conversations (c)

3.3 Collaborative activity for Libra

The ideal activity to use Libra with provides opportunities for rich conversations between students. The chosen task should involve collaborative problem solving competencies such as establishing and maintaining shared understanding, taking appropriate action to solve the problem, and establishing and maintaining team organization (OECD, 2017). The coach chose a complex skillbuilder from Lotan (2002) to serve as an example of a groupworthy task for the advisors. The chosen activity entitled Rainbow Logic is a turn-based puzzle game. The object is to discover the hidden grid in as few questions as possible. Players take turns as the 'grid designer', designing a hidden three-by-three grid of squares with three evenly distributed colors. The rules are: all squares of the same color must be connected by at least one full side; ask for the colors in a specific row or column; and colors may or may not be given in order. See Figure 6 for examples of possible and impossible grids.

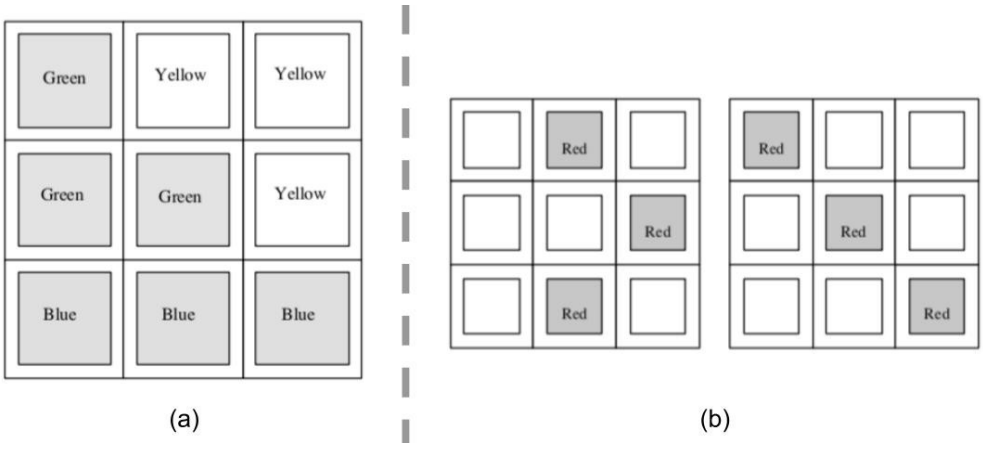


Figure 6: Possible (a) and impossible (b) grids for the collaborative activity *Rainbow Logic*

3.4 Moderation discussions

Rust et al (2005) state, “Achieving meaningful understanding of assessment requires some kind of active engagement with the criteria by both tutors and students.” To conduct moderation discussions, facilitators need to have samples of student work along with ratings against scoring criteria assessed in the work. Participants are provided a collection of scored student work and are asked to check the ratings. They respond to prompts such as, "Do you agree or disagree with this score for the student work? Justify your response with evidence from the student work and the scoring criteria." The resulting conversations are used to update the scoring criteria and build a common understanding of the grading rubric.

3.5 Teacher advisory responsibilities

Prospective advisors were provided the overview timeline in Table 2 with the goals of implementing the game with their students and then delivering a content-aligned lesson using the chat system and communication visualization introduced during gameplay. Notice that jargon to describe the building blocks of the IBDS was not used in the initial communication with teachers. The focus at this point was that teachers understood there were two implementation deliverables highlighted in grey, the game and the content lesson. Additionally, the term ‘communication viz’ was introduced to provide a name for the collaboration assessment system. Developing a shared understanding of the IBDS was an ongoing negotiation during the PD process detailed in the Design section.

Table 2: Initial teacher advisory overview timeline spring 2020

DATES	DELIVERABLE
Feb - May	Implement game with students
Feb - May	Contribute to design of communication viz
Mar - May	Design 1-2 lessons with communication viz
May or Fall	Implement content lesson with students
May or Fall	Iterate lesson with communication viz once

4. Design

The goal of the Integrated BEAR Design System (IBDS) was to establish shared understanding of each building block of the BEAR Assessment System (BAS) through iterations of the construct map, items design, outcome space, and measurement model while developing game-assisted assessments for collaboration. This section focuses on how teachers engaged in participatory design structures that gradually introduced the IBDS. Table 3 is the detailed timeline provided to teachers when they were invited to join the advisory (revised due to COVID-19). The names of these meetings and deliverables came from common teacher language around lesson planning for there to be no confusion with the language in this table and teachers took no issues with it.

Table 3: Teacher advisory detailed timeline

DATE	GOAL	TIME	WHO	LOCATION
Feb 9	Form Teacher Advisory	1 hr	SNHU:Evan TA: Each individual	Remote individually
Late Feb	TA orientation	1 hr	SNHU:Evan TA: Each individual	Remote whole-group meeting
Late Feb	Lesson plan review	2 hr	SNHU:Evan TA: Each individual	Remote individually
April	Design pre-work	1 hr	TA: Each individual	Remote individually
Late Apr	Design session	1 hr	SNHU:Evan TA: Each content team	Remote content-team meeting
Apr - Oct	Game pre-brief	45 min	SNHU:Evan TA: Each content team	Remote individual meeting
Apr - Oct	Implement game	3 hr prep 90 min	SNHU:Evan observe TA: Each individual	Remote individually
DATE	**GOAL**	**TIME**	**WHO**	**LOCATION**
Apr - Oct	Game post-brief	30 min	SNHU:Evan TA: Each content team	Remote individual meeting
Late Jun	Game Session	1 hr	SNHU: Evan, TA: Each individual	Remote whole-group meeting
Late July	Moderation Session	75 min	SNHU: Evan, TA: Each individual	Remote whole-group meeting
Late Aug	Chat Pre-work	1 hr	TA: Each individual	Remote individually
Early Sep	Chat Session	1 hr	SNHU: Evan TA: Each content team	Remote whole-group meeting
Oct	Lesson pre-brief	45 min	SNHU:Evan TA: Each content team	Remote individual meeting
Oct - Feb	Implement lesson	3 hr prep 90 min	SNHU:Evan observe TA: Each individual	Remote individually
Oct - Feb	Lesson post-brief	30 min	SNHU:Evan TA: Each content team	Remote individual meeting
Feb - Mar	Student Survey	15 min	TA: Each individual	Remote individually
Feb - Mar	Teacher Survey	15 min	TA: Each individual	Remote individually

4.1 Participatory design framework dimensions

The PD framework dimensions described in Section 2.1: form, purpose and context were communicated with the team in Table 3. The form of the activities in shaded rows was enactment because they involved implementation with students. The form of the activities in the first three unshaded rows was telling when teachers were new to the advisory and the form of the activities in the unshaded rows in April was making because they were design sessions. The form of the four unshaded rows in June – September was enactment because teachers experienced the game, moderation and chat sessions as students. The activity purpose was suggested in the Goal column by words like orientation, review, pre-brief, implement, post-brief, and pre-work. The activity context included group composition in the Who column, remote and online in the Location column, and time commitment from the Date and Time columns. What was not made explicit in the table was whether probing, priming, understanding current experience, or generating ideas about the future were the purpose of each activity. How these purposes map to each activity is described in more detail in the BAS building block sections below.

4.2 IBDS structure

The main participatory design activities in this instance of the IBDS are displayed by group context over time in Figure 7. The first whole- and small-group meetings laid the foundation for the work. It was made explicit that teachers have agency in this design process. The game and chat system were in early development during these first meetings and teachers were encouraged to provide feedback and ideas. They were not provided with scripted instructional materials to follow but rather were expected to develop the materials that would work for their students. These meetings involved individual pre-work and follow-up to set the precedent that the work would involve planning, implementation and reflection.

Each class implementation was an opportunity to provide formative assessment of collaboration skills. To ensure successful implementation, four distinct actions were taken. First, the whole group met to work through a demo lesson together. Then the coach met with teachers individually three times (shown as circles in Figure 7): first to pre-brief their lesson (pre), second to observe it (obs), and third to debrief it (de). During these coaching conversations they worked on tailoring instruction and assessment to each teacher's local needs.

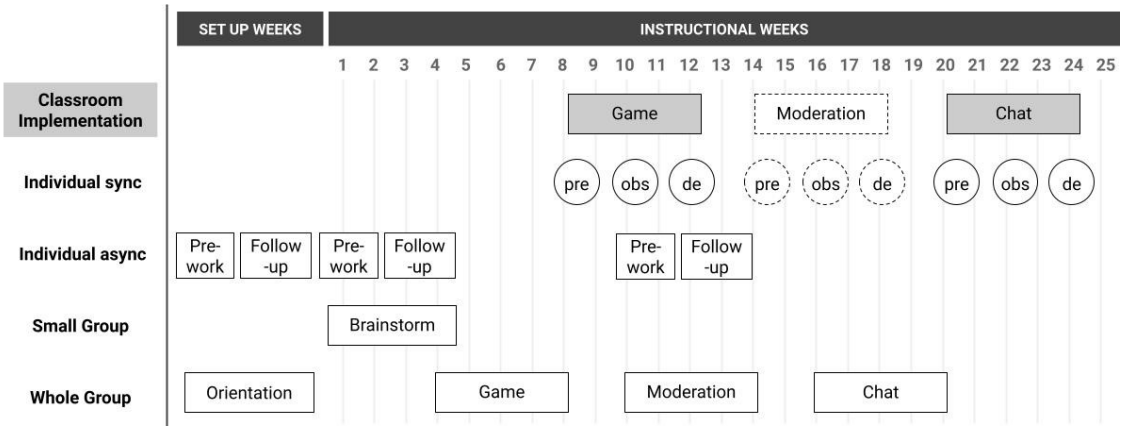


Figure 7: IBDS calendar of participatory design activities by group context

4.3 Construct mapping

The construct map answers the question, what are we teaching and assessing? The teacher advisory began this conversation as a whole group at orientation to prime teachers on the work of the advisory and introduce the guiding question, "How do we communicate effectively in teams?" A draft lesson plan was shared with teachers and the game was demonstrated for them. After orientation, teachers were given a series of follow-up tasks as pre-work for the first lesson plan session. These included sharing their current assessment practices for content and collaboration as well as providing feedback on the draft lesson plans and playable prototype of the game. The purpose of these initial activities was to prime teachers by exposing them to available resources, to probe teachers in terms of their facility with lesson planning and construct mapping, to understand current experience by getting teachers to reveal some of the realities at their sites, and to generate ideas about the future by asking teachers to anticipate how they plan to use the game and communication visualization with students.

Pairs of teacher-advisors met in content teams for the design sessions. These sessions were done in small groups to maximize output from the brainstorm process and to develop working relationships within the content teams. During these brainstorms, teachers worked on a shared Google sketch to write possible evaluation criteria for collaboration. To help describe what that meant the coach rephrased the prompt as, "What does collaboration look like? What evidence shows that students are collaborating?" Figure 8 shows the list that the Biology content team produced. The purpose of these brainstorm sessions was to understand current experience by getting teachers to list what matters most to them and generate ideas about the future by coming to consensus and agreement on the criteria that best map the construct. The coach took the opportunity after the brainstorm to share the PISA 2015 descriptors for Collaborative Problem Solving (CPS; OECD, 2017) shown in Table 4 to prime teachers with established criteria and generate ideas about the future as they made connections between their lists and PISA's CPS matrix. These lists were a source of truth during subsequent meetings to generate assessment items that elicited the expected evidence.

The construct map evolved as teacher-advisors came to consensus around the construct of collaboration within their community of practice. For example, each teacher developed a process of learning (POL) tracker from their criteria lists and by the time the coach facilitated the game session he used a remixed tracker that had gone through multiple iterations as described in the items design section. This choice focused the group to attend to specific criteria within the construct map. The purpose of focusing the teachers on the criteria with the most agreement was to probe which qualities they deemed most relevant and what ordering of qualities they anticipated as students develop their collaboration skills.

Figure 8: Google sketch notes from the biology content team design session

Table 4: Matrix of collaborative problem-solving skills for PISA 2015

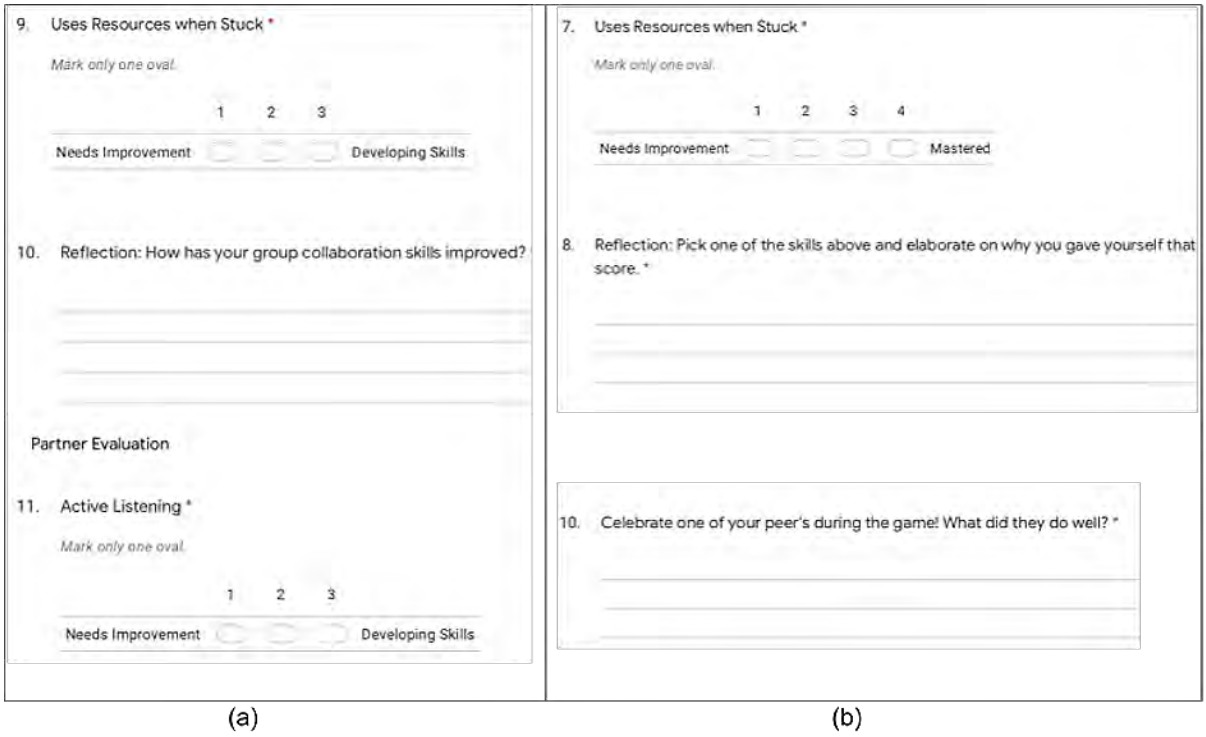
	Establishing and maintaining shared understanding	**Taking appropriate action to solve the problem**	**Establishing and maintaining team organisation**
Exploring and understanding	Discovering perspectives and abilities of team members	Discovering the type of collaborative interaction to solve the problem, along with goals	Understanding roles to solve the problem
Representing and formulating	Building a shared representation and negotiating the meaning of the problem (common ground)	Identifying and describing tasks to be completed	Describe roles and team organisation (communication protocol/rules of engagement)
Planning and executing	Communicating team members about the actions to be/being performed	Enacting plans	Following rules of engagement (e.g. prompting other team members to perform their tasks)
Monitoring and reflecting	Monitoring and repairing the shared understanding	Monitoring results of actions and evaluating success in solving the problem	Monitoring, providing feedback and adapting the team organisation and roles

4.4 Items Design

The items design answers the question, how are we teaching and assessing the construct? The draft lesson plans contained suggested activities like a think-ink-pair-share around the guiding question, “How do we communicate effectively in teams?” to generate a class list of criteria for collaboration; reflective prompts like notice and wonder; and self and peer feedback around the class criteria. The purpose of sharing these activities was to prime teachers with tasks that emphasize collaboration skills and inspire the generation of remixed materials. Teacher reactions to the lesson plans revealed how some teachers wanted to leverage the game for content and others for collaboration. One teacher suggested, “introduce tragedy of the commons and sustainability during the communications lesson” to connect it with content, while another suggested an

open-ended task, “After students play the first round of the game silent, I want them to all gather together and talk to each other for a 3-minute debrief where they can share all of their emotions: celebrations, frustrations, etc.”

All teacher-advisors used the guiding question to generate a class list of criteria during game implementation whereas the reflective prompts varied and evolved over time. Each teacher designed a unique POL tracker after the initial design session. Teacher work products were summarized and shared via email by the coach to prime teachers with more examples and inspire the generation of more ideas. One tracker that was particularly well received was a Google form co-developed by a biology teacher, math teacher and the coach. The initial version of this tracker was a Google doc by the math teacher designed to be printed that targeted seven criteria for self-rating on a 4-point scale (Beginning, Developing, Meets the standard, Exceeds the standard) and asked broad reflection questions, “What did you notice? What went well? What didn’t?” The first remix displayed in Figure 9a by the biology teacher was a Google form with a 3-point Likert scale from Needs Improvement to Developing Skills that targeted six criteria for self- and peer-rating, with a slightly more targeted reflection prompt, “How have your group collaboration skills improved?” The math teacher subsequently converted it to the second remix shown in Figure 9b. A 4-point Likert scale from Needs Improvement to Mastered with a targeted reflection question, “Pick one of the skills above and elaborate on why you gave yourself that score”, that asked students to celebrate their peers rather than evaluate them.



(a) (b)

Figure 9: An earlier version (a) and later version (b) of a remixed POL tracker

Before, during and after implementation of the game and the chat lesson each teacher-advisor was consulted on their learning objectives by the coach with questions like, “What are the learning targets?” “How are students going to demonstrate them?” and “How will progress be monitored?" The purpose of this questioning was to probe their current preparation and provide time to revise their tasks to better align with their learning objectives during pre-brief; to probe student understanding of the objectives during implementation; and to probe how successful teachers felt about meeting their objectives, gauge the current state of the class’ progress with the construct, and generate ideas for future instruction and assessment during debrief. Delivery of assessment items teachers generated to probe student understanding of gameplay and their reflections on collaboration varied across sites with a mix of Google forms and sheets, Peardeck, Zoom and Meet chat, Slack, silent signals and coming off of mute to respond.

The last whole group meeting was the content-aligned chat session, referred to as the ‘chat lesson’. The coach facilitated the Rainbow Logic activity described in the Methods section using Libra for group communication.

All participants were muted in the main room of a video call to reduce the chance of communicating through means other than Libra. The coach modeled the activity and teachers took turns in groups of three as the 'grid designer'. After multiple rounds, the coach then facilitated a discussion while displaying the communication visualizations captured by Libra. The purpose of this discussion was to model how the data might be leveraged in a virtual classroom, have teachers reflect on the realities of using this tool with their students, and discuss talk moves and purposeful questions that could lead to generative classroom conversations. The visualizations showed that one person in each group was dominating the messaging as displayed in Figure 10. This led to discussions around why certain players sent more or less messages than others, and teachers saw value in frequency counts to monitor participation. However, aside from being a useful management metric, "quantity doesn't get at quality" as one teacher put it. The recommendation from the advisors was to find instances in the chat log that demonstrate qualities of collaboration to share with the class, and to discuss types of contributions that have a large impact on moving the group forward.

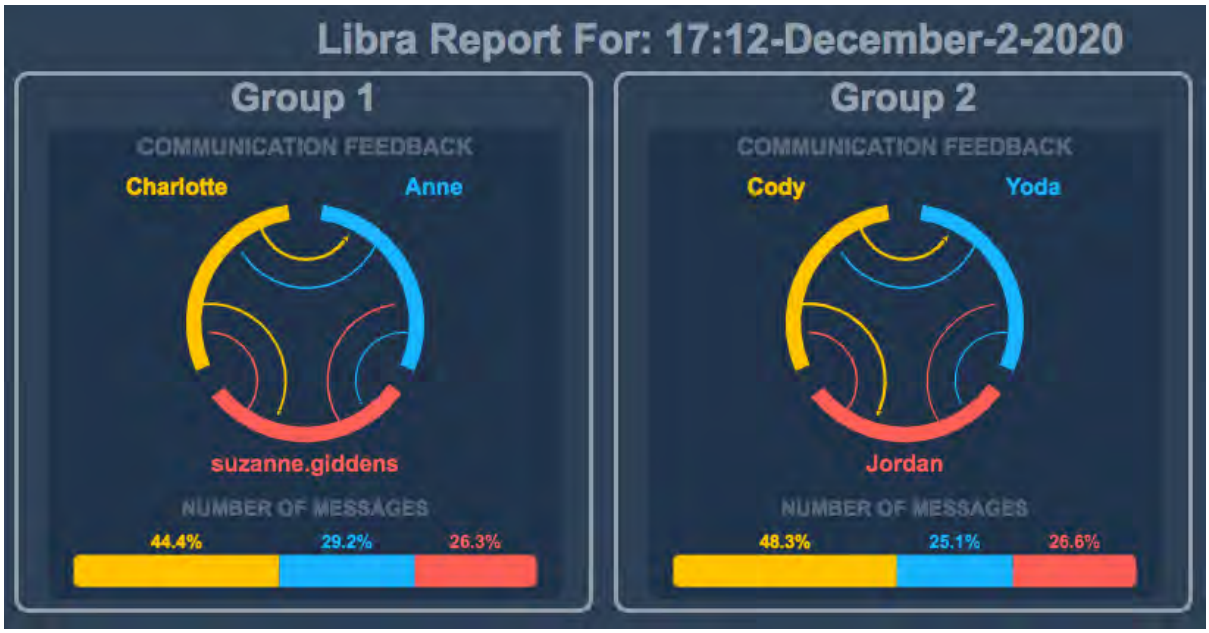


Figure 10: Libra report for the chat lesson with teacher-advisors

We discussed how each teacher planned to use Libra in their unique environments. Ideas generated from that discussion included: playing battleship with linear inequalities, sharing code to draw a snowman, and using Libra as a backchannel when working on content-related group slides. Overall, teachers were excited about the versatility of Libra. The chat lessons allowed them to use their own instructional activities with content-aligned learning targets. The construct map and outcome space supported ongoing formative assessment of collaboration in addition to and not instead of the primary course content.

4.5 Outcome space

The outcome space answers the question, what was learned? It specifies the evidence that demonstrates understanding for various levels of sophistication within the construct map. A backwards-planning process was followed whereby the outcome space was drafted before tasks and prompts were made for students. In the pre-work for the first design session teachers were asked, "What does assessment and feedback for collaboration look like in your class?" The purpose of this activity was to probe teachers' assessment practices, understand the current classroom experience of collaboration, and prime the group to have an in-depth conversation. Responses to that question surfaced one teacher's 4Cs rubric, another's POL tracker and gave the coach an opportunity to describe the PISA 2015 CPS matrix.

After brainstorming evidence statements to initially map the construct, each teacher-advisor was given the task of developing a POL tracker for their specific context and to respond to the following prompt, "My question for everyone (including myself) is how does another teacher use your tool to score student performance? What evidence do they look for to differentiate between a score of 2 and 3? 3 and 4?" The purpose of this task was to prime teachers to develop an outcome space and measurement model, to probe

teachers' ability in assessment modeling, to understand current experience by getting teachers to describe how these process of learning trackers might work in their context, and to put teachers on the spot to generate tools that they would actually use in their classrooms to assess collaboration. POL trackers were based on content team evidence statements and classroom needs. Notice in Table 5 that all of these trackers contained three goals that describe similar qualities of communication and collaboration as the three CPS competencies of PISA 2015. This alignment allows the individual teachers to use their personalized POL trackers for their specific context, while still tracking evidence that maps to established criteria.

Table 5: Process of learning tracker goals and CPS competencies of PISA 2015

MATHEMATICS	**BIOLOGY**	**BIO / CS**	**PISA 2015**
Shares Thoughts/Plans/Ideas	Team Communication	Share and listen to ideas/Ask questions	Establishing and maintaining shared understanding
Consistent effort	On Task	Work productively towards the task	Taking appropriate action to solve the problem
Peer Interaction	Team Management	Cooperate	Establishing and maintaining team organisation

4.6 Measurement model

The measurement model answers the question, what do the scores mean? The IBDS allows teachers to incorporate elements from their existing assessments and determine a measurement model that makes sense for the target construct with their students. Responses to the POL tracker follow-up prompt, "how does another teacher use your tool to score student performance? What evidence do they look for to differentiate between a score of 2 and 3? 3 and 4?" revealed the teacher-advisors' initial thoughts on a measurement model. They all provided a scale and described how a student might get a particular score. Two teachers used teacher evaluation, four used student self-evaluation and two of those four also used peer evaluation. More than half of the teachers mentioned how subjective this process of assessing "soft skills'' was and one stated that tuning was an important part of the process with rubrics, which was a perfect segue to introduce the second design session.

The second design session was a moderation discussion to engage with the collaboration criteria. Teacher-advisors were provided with short summaries of their game sessions as pre-work to remind them of their game experience. This session was done as a whole group so that teachers from all content areas could develop a shared understanding of the outcome space and consider possible measurement models. Teachers had completed a POL tracker after gameplay to self-assess on collaboration criteria and the Little Fish Lagoon teacher dashboard captured transcripts of group conversations. The coach compiled the results from the self-evaluations along with the chat logs to use as evidence to justify ratings. The spreadsheet illustrated in Figure 11 was the main referent during the moderation discussion. The purpose of this activity was to probe whether everyone was interpreting scores in the same way, to prime teachers to conduct moderation discussions with their students, and to generate ideas about which criteria were best to target and how to describe, "qualitatively different levels of responses associated with the construct map" as Kennedy (2005) defines the outcome space. Teachers noticed qualities of effective teamwork that weren't adequately described by the criteria. For instance, the player shown in Figure 11 messaged game features like fish types, money, and biodiversity, which helped build shared understanding and make progress toward the goal, but the collaborative merit of those messages weren't adequately described by the evaluation criterion, *sharing ideas/explaining thinking*. This suggested a revision to the outcome space to include when a group member contributed a message that "advanced the group" toward their common goal. Another example of revised criterion was describing *active listening* as "paying attention and reflecting back". These elaborated evidence statements inform the measurement model by suggesting potential look-fors in and coding schemes for chat data.

Round	Message	Active Listening	Ask Questions of Peers	Shares Ideas/Explain Thinking	Takes a Risk	Works Persistently	Uses Resources when Stuck	Reflection: Pick one of the skills above and elaborate on why you gave yourself that score.
1	are you both in the game?	2	3	3	3	4	3	I gave myself a 3 (proficient) on Shares Ideas / Explains thinking because I communicated often and effectively in the chat. I shared what I was doing and asked questions of my peers in order to help them understand how the game worked and to figure out how to maintain biodiversity.
1	emote:(THUMBS-UP)							
1	make sure to fish BEFORE you click on the button "READY"							
2	looks like are biodiversity is pretty good							
2	i don't know how many fish to leave I think it's learn by trial and error							
2	I clicked one boat on Sardines and one boat on Haddock							
3	I wonder if we should click on the Research Lab to look at our data							
3	it looks like we're still doing well with biodiversity!!							

Figure 11: Spreadsheet of chat logs with self-reflections used for the moderation discussion

There was consensus that it was difficult to measure many of these qualities. As one teacher put it, "it's hard to quantify what is qualitative". For example, the criterion working persistently was difficult to gauge. Participating a lot in the chat didn't necessarily indicate this. It would require additional hooks to catch when students made mistakes and learned from them or identified obstacles and overcame them. Similarly, active listening was hard to identify when only looking at one person's part of the chat. It would have been easier to find responses that indicate they were engaging in listening to what the other players were saying if the coach had provided each group's chat log as a single transcript rather than each player's individual transcript. Even still, would tallying up the number of instances that a player demonstrated a skill be an accurate proxy for their skill level? For instance, it was suggested that asking questions could be measured by using ctrl-f to find the number of "?" characters in a transcript, but this gave people pause. Soon after, the same teacher who had made the suggestion said of the data, "it might be more useful for students to engage in and find their own value in, rather than us as teachers trying to mine insights from."

5. Discussion

Game-based assessments of deeper learning competencies have great promise for broad adoption. The Integrated BEAR Design System (IBDS) demonstrated that teachers from varying contexts could build common criteria and formative assessment systems for collaboration skills. The design was such that a game was used to introduce students to a chat system and communication visualization that could subsequently be used during content-aligned lessons to provide evidence to support claims about collaboration skills. In this section we discuss the benefits of this model, implications for other designers, areas for improvement, and future research.

5.1 Broad adoption

To broaden adoption of this game-assisted formative assessment of collaborative problem-solving skills, six teacher-advisers from diverse learning environments co-designed construct maps, tasks, outcome spaces, and measurement models to tailor assessment to their contexts and integrate it with their instruction. The resulting outcome spaces were surprisingly similar and this is encouraging for the future of game-based assessment of deeper learning constructs. If it remains true that cross-cutting themes and skills like collaboration or problem solving are defined similarly across contexts, the implication for games that target them would be larger market penetration due to the usefulness across disciplines and grade levels. Additionally, the game and chat system serve as a model of game-based scaffolding to introduce a communication tool that can be used across time within a curriculum to support ongoing formative assessment. Finally, this instance of the IBDS provides an example of game-assisted formative assessment, where the assessment is based on data that can be collected both inside and outside of the game. This provides opportunities for formative assessment of collaboration skills during lessons that target primary course content.

5.2 Participatory design

Developing assessments for deeper learning is an ongoing process. It wasn't until mid April that the coach began using the term construct map and outcome space to describe the work started in February. If

participants had been told they were going to have *moderation discussions* about *evidence statements* for collaboration skills during orientation, there would have not been enough buy-in and collective understanding for those words to have shared meaning. Active engagement must be monitored throughout the design process and academic jargon and frameworks should be gradually introduced as relationships develop. It was critical that design started by understanding what the teacher-advisors were already doing in their classrooms to teach and assess collaboration. It is tempting to skip steps in the process. To reiterate a piece of advice from Rust et al (2005), "simply being given a model answer, or a marking guide, or a set of criteria by the course or module leader will not in itself ensure a common informed understanding." Repeated cycles of making and enacting, followed by more making and enacting, is required to truly develop shared understanding.

5.3 Motivations for iteration

There are multiple benefits from repeating Participatory Design (PD) actions and the building blocks of the Bear Assessment System (BAS) with participants. The BAS is an iterative process. As a construct map becomes more sophisticated it begins to resemble a learning trajectory as described in Wilson (2009) as an "ordering of qualitatively different levels of performance focusing on one characteristic." Each pass through the building blocks improves understanding of the construct, elaborates evaluation criteria, refines items and is an opportunity to consider the validity and reliability of the measurement model.

Repeated PD actions and structures of engagement familiarized teachers with what was expected of them and provided multiple opportunities for them to build upon their ideas around assessment and instruction. Norms for interaction were established including the structure of meetings; modes of communication; and the plan, implement, and reflect cycles of our making and enacting.

The IBDS brought together a diverse group of educators who are able to share strategies, instructional materials, and creative approaches to implement the technology tools and support student development of collaboration skills. Teachers met for design sessions, engaged in the game lesson as a group, implemented the game lesson in their classes, engaged in a moderation discussion as a group, engaged in a chat lesson as a group, and implemented the chat lesson in their classes. Each activity provided another opportunity to map the construct, design items to elicit evidence of student understanding, discuss the outcome space, and to work on a measurement model that would support student learning of collaboration skills. These opportunities naturally arose from repeated questions like, "What are the learning goals of the lesson? What evidence will students produce to demonstrate understanding? How will you check for understanding? How will you determine a student's level of understanding?"

5.4 Areas for improvement

One improvement to be made is to streamline moderation discussions for classroom use. The evolution of the moderation discussion occurred during the design process and wasn't anticipated during the early development stages of the game and chat system. Currently, to acquire the chat message log teachers had to copy the raw text of the chat log from the webpage or be sent a table from collected telemetry data with each message event. To improve this we can automate the creation of a chat log table.

Another improvement is to include a POL tracker with both Little Fish Lagoon and Libra so that students don't have to go to an external form to reflect on their collaboration. Ideally, this feature would allow for custom trackers so teachers could tailor the evaluation criteria as they did while participating in the IBDS. This would enable us to provide data displays similar to those used by the coach during the teachers' moderation discussion, to collect and display reflection data over time, and scale-up the work of iterating on the building blocks of the BAS for collaboration skills.

5.5 Assessing complex domains

A simple rule-based approach to assessment may not work for the complexity of this domain, and perhaps a teacher mining for insights is also untenable, but that doesn't rule out machine mining for insights. Behrens, DiCerbo and Foltz (2019) describe assessment advances in natural language and collaborative problem-solving using machine learning-based analyses of data. The teachers agreed that there was value in the discussion to build self-awareness in their students and help to establish and maintain norms for collaboration. If an assessment system could automatically categorize players into different types of collaborators based on their messages in a manner interpretable by teachers and students, it could result in more of these valuable

discussions. For example, training a model to identify types of contributions that have a large impact on moving the group forward or that identify instances of active listening could provide snapshots for teachers to orchestrate conversations around. This is an interesting direction for future research.

The communication visualization tool that is embedded in both Little Fish Lagoon and Libra does not judge student performance. Currently, it is a data collection tool that displays chat logs and frequency of messages sent. These are examples of unobtrusive data streams that are enabled by interactive applications on network-connected devices. The IBDS involves teachers in the process of deciding how to leverage these data in a way that is integrated with the process of learning. The value proposition for automated judgments of students' collaboration skills based on these data is still unknown. It might be more useful for teachers to engage in and find their own value in, rather than machines trying to mine insights from.

Acknowledgements

The William and Flora Hewlett Foundation supported this project. We'd like to thank the game team at SNHU Innovation Center, San Francisco. In particular we would like to acknowledge the work of Jordan Suhr, Ted Southard, Steve Swink, Andrew Dakhil, Keisha Sheedy, and Holly Holtz in the design and development of the game Little Fish Lagoon and the standalone communication visualization tool, Libra. Additionally, we acknowledge the Teacher Advisory recruited for this project including Cody Holland, Charlotte Sivanich, Sophia Sheena, Anne Rizzacasa, Lauren Schulz, and Suzanne Giddens – all of whom made substantial contributions to the instructional materials and embedded assessment system.

References

BEAR Center, 2014. The BEAR assessment system, [online], University of California, Berkeley, Available at <https://bearcenter.berkeley.edu/page/bear-assessment-system> [Accessed 11 Jan 2021].

Behrens J., DiCerbo K., and Foltz P., 2019. Assessment of complex performances in digital environments, *The ANNALS of the American Academy of Political and Social Science*, Trento, Italy, 683(1), pp 217-232. doi: 10.1177/0002716219846850.

Brandt E., 2006. Designing exploratory design games, *PDC: Expanding Boundaries in Design*, Trento, Italy, 1(1), pp 57-66.

Brandt E., Binder T., and Sanders E., 2012. Tools and techniques: Ways to engage telling, making and enacting. In: Simonsen J. ed., and Robertson T., ed. *Routledge International Handbook of Participatory Design*. Routledge International Handbooks, pp 145-181.

Drain A. and Sanders E., 2019. A collaboration system model for planning and evaluating participatory design projects, *International Journal of Design*, [Online] 13(3). Available at <http://www.ijdesign.org/index.php/IJDesign/article/view/3486> [Accessed 08 Jan 2021]

Graesser A. C., Foltz P.W., Rosen Y., Shaffer D.W., Forsyth C., Germany ML., 2018. Challenges of Assessing Collaborative Problem Solving. In: Care E. ed., Griffin P. ed, and Wilson M. ed. *Assessment and Teaching of 21st Century Skills: Research and Applications*. New York, NY: Springer, Cham, pp 75-91. https://doi.org/10.1007/978-3-319-65368-6_5

Hamari J. and Nousiainen T., 2015. Why Do Teachers Use Game-Based Learning Technologies? The Role of Individual and Institutional ICT Readiness. *2015 48th Hawaii International Conference on System Sciences*, Kauai, HI, USA, pp 682-691.

Jean Justice, L., and Ritzhaupt, A. D., 2015. Identifying the barriers to games and simulations in education: Creating a valid and reliable survey. *Journal of Educational Technology Systems*, 44(1), pp 86-125.

Kennedy, C., 2005. "The BEAR Assessment System: A Brief Summary for the Classroom Context", University of California, Berkeley.

Lotan, R.A., Bunch G. and Gainsburg J., 2002. Complex Skillbuilders [Online]. Available at: <https://web.stanford.edu/class/ed284/csb> [Accessed: 11 Jan 2021]

National Research Council. 2012. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Committee on Defining Deeper Learning and 21st Century Skills, J.W. Pellegrino and M.L. Hilton, Editors. Board on Testing and Assessment and Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nousiainen, T., Kangas, M., Rikala, J., & Vesisenaho, M. (2018). Teacher competencies in game-based pedagogy. *Teaching and Teacher Education*, 74(1), pp 85-97.

OECD, 2017. *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving,* PISA, OECD Publishing, Paris, pp 131-150.

Rust C., O'Donovan B. and Price M., 2005. A social constructivist assessment process model: how the research literature shows us this could be best practice, *Assessment & Evaluation in Higher Education*, 30(3), pp 231-240.

Sánchez Mena, A. A., & Martí Parreño, J., 2017. Drivers and barriers to adopting gamification: teachers' perspectives, *Electronic Journal of e-Learning*, 15(5), pp 434-443.

Sanders E., Brandt E. and Binder T., 2010. A Framework for Organizing the Tools and Techniques of Participatory Design, *PDC '10: Proceedings of the 11th Biennial Participatory Design Conference*, Sydney, Australia, pp 195-198.

Stieler-Hunt, C. J., & Jones, C. M., 2017. Feeling alienated–teachers using immersive digital games in classrooms. *Technology, Pedagogy and Education*, 26(4), pp 457-470.

Stieler-Hunt, C., & Jones, C., 2019. A professional development model to facilitate teacher adoption of interactive, immersive digital games for classroom learning, *British Journal of Educational Technology*, 50(1), pp 264-279.

Watson, W. & Yang, S., 2016. Games in Schools: Teachers' Perceptions of Barriers to Game-based Learning. *Journal of Interactive Learning Research*, 27(2), pp 153-170.

Wilson, M., 2009. Measuring progressions: assessment structures underlying a learning progression, *Journal of Research in Science Teaching*, 46(4), pp 716-730.

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

Yuan, Kun and Vi-Nhuan Le, 2014. *Measuring Deeper Learning Through Cognitively Demanding Test Items: Results from the Analysis of Six National and International Exams*, Santa Monica, CA: RAND Corporation, https://www.rand.org/pubs/research_reports/RR483.html.

Appendix

Table 6: School demographics and descriptions

Grades	Overall rank	Students per grade	Minority	Subject area	Description
9-12	top 20	121	79	Math	Small, high performing, urban, diverse public school in Western U.S.
K-12	top 20	140	68	Math	Small, high performing, course-based independent study charter school in Western U.S. with 5 hours of instruction per course per week that meet twice a week as a group and individually 2-3 a week in the learning centre.
9-12	bottom 50	350	34	Biology	Medium, low performing, urban, homogenous public school in North Eastern U.S.
9-12	bottom 50	200	43	Biology	Small, low performing, urban, slightly homogenous public school in North Eastern U.S.
9-12	bottom 50	403	82	Biology	Medium, low performing, rural, diverse public school with a 21st century skill focus in Western U.S.
7-8	top 30	880	46	Game & Code	Large, urban, high performing, slightly homogenous public school in Central U.S.

Quantitative data from https://www.publicschoolreview.com/[[[]