



Better Learning in Games

A Balanced Design Lens
for a New Generation of
Learning Games

LEARNING
games
NETWORK



Massachusetts
Institute of
Technology

Scheller Teacher Education Program
the education arcade

Created by

Jennifer Groff, Ed.M.
Jody Clarke-Midura, Ed.D.
V. Elizabeth Owen, Ph.D.
Louisa Rosenheck, Ed.M.
Michael Beall

With contributions from:

Eric Klopfer, Ph.D.
Mike Tissenbaum, Ph.D.
Matthew Berland, Ph.D.

This guide is presented by designers, researchers, and developers at the



and



Scheller Teacher Education Program
the education arcade

—whose mission is to support the advancement of the field of learning games.

About

This is a guide to inform next generation learning game designs. It is an introduction to the core approaches used by many learning game designers, including the Learning Games Network and The Education Arcade at MIT.

The aim of this guide is not to impose a new game development method—but rather to present a learning design approach to be integrated with current best practices in game design.

Based on the fundamental structures of Evidence-Centered Design (ECD), we explore a design lens known as Balanced Design. The emphasis of this framework is not for building assessments into games but rather helping game designers to better align the learning goals with game mechanics to produce more deeply engaging and effective learning game experiences.

This is not meant to be a comprehensive overview of ECD or assessment-based game design, but rather an introduction to an ECD modified framework—intended to help all levels of learning game designers become familiar with a more robust model of learning game design and a powerful tool for creating the next generation of learning games. As an introduction, this guide offers the basic framework along with tools and supports for getting started. Included are several case studies that describe how this framework has been used, and begin to illuminate the further directions you can go with such a framework. Subsequent guides and papers will go into further detail on the game design models by the organizations presented here.

Who this guide is for...

... **game designers, game developers, instructional designers, educators, publishers**—or anyone interested in understanding more about learning-aligned frameworks for better game and learning experience design.

Game Art: Lure of the Labyrinth



This way...



Introduction

Learning games are not just a genre of games, but a unique and emerging field that operates at the intersection of game designers, learning designers, subject matter experts, developers, educators and researchers, who collaborate to produce innovative and powerfully engaging learning experiences. In the past 10 years, the field of learning games has grown dramatically. In that time, it has had many successes and more than our share of failures.

The past decade has been spent fleshing out the possibility space, with examples of what's possible in learning games across the spectrum—from small apps to commercial 'mods'¹ to 3D worlds, and many things in between. Yet as a field, we are just beginning to crest the horizon on what's truly possible for leveraging learning games as deeper and formative learning experiences both in and out of school.

Moving the field forward in dramatic ways, to create games that produce deeper learning but also support and guide the learner towards richer understanding and concept mastery, will require the field to take a big step forward. What tools, structures and methods will help us do that?

Achieving this new plateau will require better articulation, design, and evidence of learning goals and performance within learning games. Teachers, parents, and learners all need — and often request of game designers — a better understanding of what a game is targeting, how that situates into a larger learning sequence, and how they know what they've learned from the game.² Answering these needs is not just for the benefit of teachers and parents, but in doing so will directly benefit the learner and make a more powerful learning game. To do this, it will require a more robust learning framework at the core of the game design. Many have argued that assessment design is a key mechanism that will propel the field forward;³ yet there still exists a large chasm between general assessment design — and even game designers leveraging rigorous assessment methodologies — and the rest of the field.



At the same time, many of today's schools have an assessment-centered culture, even though assessment methodologies and the static nature of assessments often used in schools fall short in many ways—including being able to capture and measure all that is valued in education. Assessment and accountability are primary drivers in education, despite the fact that our field of assessments still has a very long way to go.

In short, *games need assessment, but assessment also needs games*. Two fields that once felt diametrically opposed, it is becoming clearer are coming to see the increasing overlap and value in how they intersect—and perhaps most importantly, that assessment need not be loathed by game designers, but might in fact be the very tool that elevates their craft to the next level. In this guide, we propose an approach and framework to help build this bridge by facilitating a more robust and coherent design of games for learning.

1 Short for 'modifications' — i.e. commercial games that have been modified for educational purposes.

2 Klopfer, E., Osterweil, S., & Salen, K. (2009). *Moving learning games forward*. A white paper by the MIT Education Arcade.

3 Phillips, V. & Popović, Z. (2012). More than child's play: Games have potential learning and assessment tools. *Kappan*, 94(2), 26-30.

An expanding genre...

The field of learning game design has made tremendous strides in the past decade. Despite this great work, there is still a long way to go before we deeply and systemically tap the power and potential of learning games. The dramatic strides of the field are still nagged by lingering questions such as,

Why aren't learning games integrated more in classrooms?

What learning is really happening in games?

How can learning games be designed to have even deeper and more meaningful impact on the learner?

Moving to the next level will require much more thoughtful and strategic mapping of the learning experiences and outputs in games. It is clear that games can be powerful learning environments, but the field needs better ways to describe how and why learning is happening in the game experience for a number of reasons—including the better identification and implementation of games by educators in the classroom, more meaningful feedback for educators on the learning experience for each student, and better data to inform the further design of educational games.

**“It’s not about assessment,
it’s about aligning the learning.”**

In just the past five years, powerful frameworks and tools for learning game design have emerged, which are helping us to begin to collect much more learning data in games, and create more responsive and formative learning experiences for learners well above and beyond the previous generation of learning games. Some of these tools are quite robust, offering game environments as authentic assessments, and are moving us in the direction of self-contained formative and responsive learning environments. Even though these tools represent the cutting edge and the *future* directions of the field, there are powerful practices built into their approach to game design which can benefit the design of all learning games *today*.

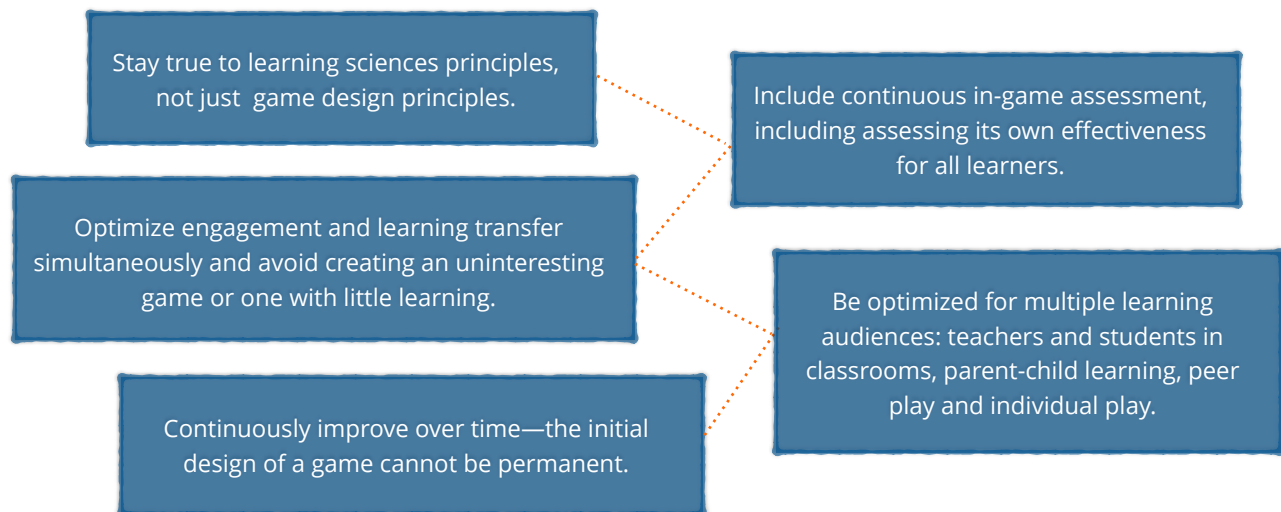
'At an Impasse': The Next Generation of Learning Games

In many ways, the learning games field is approaching its own adolescence—an inflection point, ready to take that next big step forward. We've built a wide array of games, but if learning games are to have deeper impact then they need to better demonstrate the learning that is a result of the play experience.

This adolescence is perhaps not so coincidentally aligned with an unprecedented interest, emphasis, and outright demand for innovation in assessment in education. This demand has placed quite a spotlight on games and their potential as assessment mechanisms.

But for many learning games designers, building traditional assessments is not why they got into this field and is the surest way to kill the fun of any learning game. Yet as we shall see, the underlying design frames of assessment are actually an opportunity for more meaningful learning game play.

Games and assessment structures are a unique and useful marriage, and assessments need games just as much as the field of learning games needs assessment. It's actually the intersection of these two that has the potential to move the field forward to more engaging and transformative play experiences. In discussing this potential, Vicki Phillips of the Bill and Melinda Gates Foundation, and Zoran Popović, the lead designer of the notable game *Fold.it*, offer what they believe to be a key set of design principles that learning games must adopt in order to be effective:¹



As demonstrated by their suggested design principles, the next generation of learning games puts learning — and evidence of that learning — at the core of the game's design. And when you do, the game's design, the game, the learning, and the play experience, take a big step forward.

Why Assessment?

The intersection of assessment and game design is quickly becoming a hot topic—and for good reason. Our traditional notions of assessments just won't cut it in the 21st century. But likewise, learning games need to do a better job of aligning with learning goals and standards, and be able to communicate student performance and what the learner understands in relation to those goals as a result of playing the game. This feedback from the game is critical to supporting learning for both the learner and the teacher. There are great advancements at the intersection of assessment and game design in the works right now — with more and more examples from small apps to large game worlds — being built on robust data structures.

For many learning games designers, that might not be their focus or goal. But assessment often gets a bad rap, and might not necessarily be what you think it is. Broadly speaking, 'assessment' is about creating ways to identify how you know what the learner has learned, so there is good reason to give assessment-based game design a second look. The frameworks underlying many of these game designs offer a more powerful design model for the next generation of learning games.



**“Not assessment,
but a better aligned
learning model.”**

For many learning game designers, the mention of the word “assessment” is an automatic deal-breaker—a surefire sign of certain death of a fun, well-designed and engaging learning game. And at a surface level, that can be true. Yet, it turns out, that if you go deeper, certain assessment frameworks serve as fantastic design models at the core of a learning game's design for better learning, and better play.

Although the framework discussed in this guide has its roots in assessment design, at its core it's just a framework that helps to better align the learning goals, tasks and outcomes of any learning experience. The nature of the framework allows it to serve as a very robust frame or 'spine' of a learning game's design, which can not only help the game have much a deeper impact on student learning but help educators and parents have a much better understanding of what the game is targeting and exactly how it is helping students.

**“Better learning,
and better play.”**

Frameworks for structuring learning in games

One framework that has become particularly popular amongst learning games designers is Evidence-Centered Design (or ECD), because it offers a powerful conceptual design framework that can be used to collect assessment data in many types of formats—including digital games. The ECD approach to constructing educational assessments focuses on measurable evidence of a student's learning. It was developed at the Educational Testing Service (ETS) by Robert Mislevy, Russell Almond, and Janice Lukas, and has been built upon extensively by many other researchers. Evidence-Centered Design offers a methodology for collecting and analyzing evidence from tasks performed by the student.¹

A number of learning game designers have adopted and built upon the ECD model for creating game-based assessments—including the work of Valerie Shute and her team at Florida State University on 'stealth assessments' and the work of GlassLab Games.² Much of this work represents the cutting-edge of the intersection of games and assessment. And though powerful, this model of game design can be very labor intensive. Implementing the full ECD framework for designing true psychometrically valid assessments often takes many months to develop the content model³ behind the game, before the game design work even begins.⁴

For the general learning game designer, this approach may not be necessary or even possible—and many are not interested in developing psychometrically valid assessments in their games. Yet for those designers, to throw out ECD entirely is a missed opportunity. The driving design questions and approach that lie at the heart of ECD offer a powerful lens for dramatically improving the design and opportunities in learning games. As a result, there has been a dramatic increase in the application of ECD across the learning game community.⁵

In this guide, we will discuss how a design lens based on Evidence-Centered Design we call **Balanced Design** can serve as a powerful tool and frame for the design of a learning game, as well as the impact and opportunities for the designers as well as the educators and students if this approach is used.

1 Mislevy, R., Almond, R., & Lukas, J. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, 2003(1), i-29.

2 Shute, V. J. & Ventura, M. (2013). Measuring and supporting learning in games: Stealth assessment. Cambridge, MA: The MIT Press. Mislevy, R., Oranje, A., Bauer, M., von Davier, A., Hao, J., Corrigan, S., Hoffman, E., DiCerbo, K., & John, M. (2014). Psychometric Considerations In Game-Based Assessment. A white paper by GlassLab Research and Institute of Play.

3 See figure on next page.

4 Shute, V. & Ventura, M. (2013). *Stealth Assessment Measuring and Supporting Learning in Video Games*. The John D. and Catherine T. MacArthur Foundation Reports on Digital Media and Learning. Cambridge, MA: MIT Press.

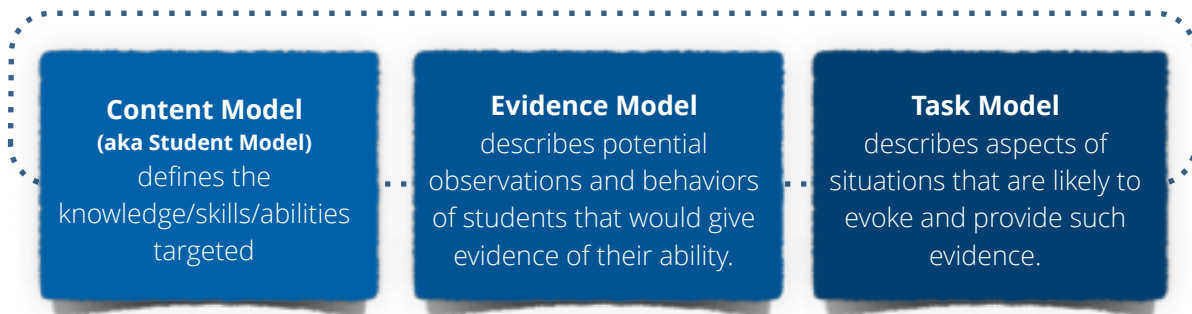
5 e.g.: Shute, V. & Torres, R. (2011). Where streams converge: Using evidence-centered design to assess Quest to Learn. In M. Mayrath, J. Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 91-124). Charlotte, NC: Information Age Publishing.; Clarke-Midura, & D. H. Robinson (Eds.), *Technology-based assessments for 21st century skills: Theoretical and practical implications from modern research* (pp. 91-124). Charlotte, NC: Information Age Publishing.; Code, J., Clarke-Midura, J., Zap, N. & Dede, C. (2012). Virtual performance assessment for serious games and virtual worlds. In H. Wang (Ed.), *Interactivity in E-Learning: Cases and Frameworks*. New York, NY: IGI Publishing.; Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *The Journal of Technology, Learning and Assessment*, 8(4); Conrad, S., Clarke-Midura, J., & Klopfer, E. (2014). A Framework for Structuring Learning Assessment in an Educational Massively Multiplayer Online Educational Game – Experiment Centered Design. *International Journal of Game-Based Learning*, 4(1), 37-59.

The basics, and heart, of the ECD framework

The full ECD framework is a powerful and robust approach for assessment and assessment-based game design. There are a number of key elements, including:

- (1) *Domain Analysis*,
- (2) *Domain Modeling*,
- (3) Developing the *Conceptual Assessment Framework* (CAF),
- (4) *Assessment Implementation*, and
- (5) *Assessment Delivery*

However, it's the CAF that offers tremendous value for all learning games because it defines and aligns the content, evidence and criteria for understanding a student's learning or performance on a task, or game. To do this, the CAF breaks this down into three key elements called models:



**When these elements are designed and aligned
as the foundation of a learning experience,
there are many benefits gained.**

Together, these elements provide a framework for,

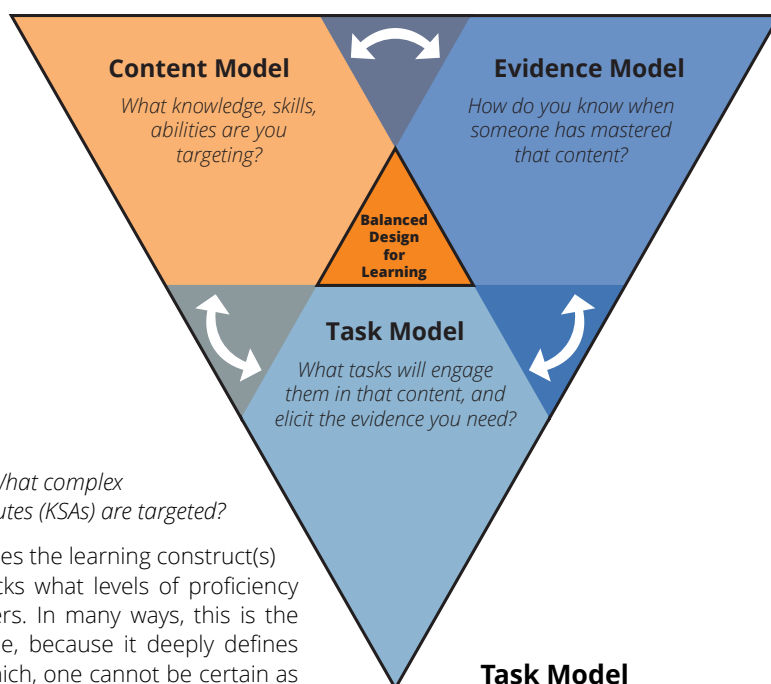
- (1) specifying the knowledge and skills to be explored (and possibly measured);
- (2) the tasks that can engage learners in regard to that knowledge and skills; *and*
- (3) the data and evidence generated from these tasks and how you will interpret it to make inferences about the learner's ability.

Put simply, together these elements create a feedback loop for an ongoing learning experience. Though this framework is a formal structure for assessment design, these elements of ECD can also be used more informally as powerful design frames for learning games and experiences .

Balanced Design

A Design Lens for Learning Games

Balanced Design offers a simple design lens for creating aligned and powerful learning experiences. This framework uses the core structural models of ECD as design lenses for the key elements of a learning game design, and when properly coordinated, creates an alignment in the learning experience design. Using a balanced design lens creates a learning game where the learning goals, game mechanics and judgements about learner play and performance are aligned.



Content Model

Answers the question, “What complex knowledge, skills, or attributes (KSAs) are targeted?”

The Content Model defines the learning construct(s) of the game, and unpacks what levels of proficiency might look like in learners. In many ways, this is the heart of a learning game, because it deeply defines the content—without which, one cannot be certain as to what the game is actually addressing. This is the element that is often too thinly addressed when designing a learning game. Elements include:

- description of the core construct
- deconstruction of the parts of that construct
- descriptions of proficiency

Task Model

Answers the question, “What are key tasks in the game, and how will the situations be structured to obtain the kinds of evidence needed?”

The task model describes the types of game features, tasks, and experiments the learner will engage in, including a step-by-step description of the game tasks or quests, as well as artifacts or data produced through game play.

Evidence Model

Answers the questions, “How will it be measured?” and “What behaviors and performances will be used?”

The evidence model defines the rules for how to interpret a learner’s performance on each task and how the content model should be updated given this interpretation (i.e. does the learner grasp the construct or do they need additional support?).

Where Balanced Design Intersects with Learning Game Design Practices

Balanced Design is less formal than ECD for a reason: it offers a more accessible framework to game designers who aren't seeking such a formal role with assessment—while still providing a powerful lens for designing the heart of a learning game. Learning games are inherently about the intersection between the playful exploration of concepts, and the artful marriage of game mechanics to the essence or nature of a concept. The Balanced Design lens presented here isn't meant to replace existing good learning game design practices—to the contrary, we still hold up the notion that playful engagement and fun must be primary objectives in designing the game and experience. But every learning game is built around a concept or topic—hence what makes it learning game—and as a result every game design team already inherently works with learning goals and fleshes out elements of the ECD framework to some degree. Balanced Design aims to extend this work, and offers a more rigorous and coherent framework that builds a better and more effective game for the learner—but also ultimately sets a better foundation for anyone to make sense of the learning game experience.



Balanced Design in Practice

Balanced Design serves as a guide for the process as well as a frame for the learning and concept elements of the game—and by its very nature, encourages and supports an iterative design process. As a result, implementing Balanced Design shouldn't be radically different from prior practices in learning game design. By using the questions in each of the three models in Balanced Design, you can better align and document the elements of the learning game.

So, when starting a new game design project, as a design team you probably already have a learning topic or concept in mind. This would be the start of fleshing out the Content Model. You probably have also thought about the game structures and mechanics that interest you.

This would be the beginning of the Task Model. The gap that's often not filled is the Evidence Model; *what data and evidence will the player generate in your game? how will you make use of it to inform further game play?* These questions demonstrate how helpful it is to know what concept mastery looks like, what may be common misconceptions or errors, and understanding levels of proficiency would be helpful. In answering these you further flesh out your Content Model, which may cause you to tweak your task and evidence model. This iterative process is executed by the design team, as game designers work in concert with instructional designers, content or subject matter experts, and at times even assessment specialists. Going through this process with Balanced Design can take some work but the benefits make it worth it, and in the end achieve many of those goals laid out by Phillips and Popović.

Case Study

Designing quests in *Radix*, an MMO for STEM

www.radixendeavor.org

The Radix Endeavor is a Massively Multiplayer Online Game (MMOG) being developed by The Education Arcade at the Massachusetts Institute of Technology, designed to improve learning and interest in STEM in high school students. The content specifically focuses on statistics, algebra, geometry, ecology, evolution, genetics, and human body systems. Players take on the role of mathematicians and scientists and embark on quests that encourage them to explore and interact with the virtual world through math and science. Players become embedded in a narrative in the world where they encounter a villain who does not believe in the practices of science. Players have to reason about science issues applicable to game characters' everyday lives, refute the unscientific claims of the villain, and make choices based on what they consider to be valid evidence.



Currently, there are seven quest lines in Biology and six in Math. Each quest line contains a series of quests around a specific content area (i.e. human body systems, geometry, etc.). The quests start out easier and get progressively harder, building off what is learned in the previous quest. The goal is to have a culminating quest activity at the end of each quest line where students will apply what they have learned to a problem or new situation.



Using the ECD framework as a guide for quest design, the Radix team then modified it to fit the project in a way that ensures the method by which evidence is gathered and interpreted in game play is consistent with the learning objectives the game is attempting to teach. What ECD describes as “work products”, the Radix team refers to as “experiments”.¹ Each quest is designed around one or more experiments that a student completes over a series of steps. For example, students may be asked to breed a particular type of flower or to build a scale map of a city. These experiments have inputs (the flower or object a student tries to breed) with particular properties. The outputs of the experiment (e.g. the flower or object that is bred), also have these properties. These data are captured and stored on the back end and can be evaluated in real time to see what kinds of experiments students are conducting—are they breeding the right objects? Are they producing the right proportion of offspring? Thus, in the quests, the experiments allow the Radix team to assess students' understanding of the content.

¹ Conrad, S., Clarke-Midura, J., & Klopfer, E. (2014). A Framework for Structuring Learning Assessment in an Educational Massively Multiplayer Online Educational Game – Experiment Centered Design. *International Journal of Game-Based Learning*, 4(1), 37-59.

This information can then be used formatively, to provide feedback to both students and teachers as a way to ensure students learn the content. It can also be used to assign “side quests” in particular content areas where students are struggling. In the culminating quests, which are essentially a performance assessment, the team will use the data to summatively assess what a student knows after working through the quest chain. Some of the variables being measured in the culminating quests will also be evaluated with external measures. This method of using evidence-centered design for the design experiment-based game design we refer to as XCD, and has allowed the team to account for the use of experiments in game play, and perhaps more critically, to make the process accessible to the various people involved (designers, content experts, programmers).



Example of game-play in relation to the XCD framework for the game RADIX

In order to articulate this process, the Radix team uses Quest Templates and tables (see example below). Each application of the template is modified as necessary through an iterative process to ensure that the team knows (1) what data to capture from the student interacting with quests, (2) if students understand the concept, *and* (3) whether or not we need to give feedback or have the student repeat the quest.



QUEST TEMPLATE

Content Model		Task Model		Evidence Model	
				Data Collected	Interpreting Evidence
Learning Objective	Quest	Task/Action			
Recognize patterns in data sets	ST1.1	Turn in data summary to support/refute government claim		Data summary (see Table 1.7 for possible data summary submissions)	<i>Correct:</i> Player knows to use a large enough sample size and the correct measure. <i>If Incorrect:</i> <i>Species other than blackburn:</i> Player likely does not know what a blackburn is <i>Trait other than body length:</i> Player did not understand what needed to be measured <i>If result does not agree with their answer in Tumbler 1.2:</i> See Quest 1 7a <i>If sample size is too small:</i> Player does not understand the importance of a large enough sample
Use models and simulations to make inferences and conclusions	EV3.3	Students use a simulator to see how environmental pressures can affect trials.		Students turn in EvoGlobe and respond to questions. Data collected includes Globe setting and responses.	<i>Correct:</i> See EvoGlobe settings EV3.3a table. <i>Potential reasons for Incorrect EvoGlobe:</i> 1. They don't understand principle of natural selection. 2. They don't understand how to interpret EvoGlobe. 3. They don't understand how to use the EvoGlobe.

Why Balanced Design?

Thus far, we've discussed how the full ECD framework can be a powerful tool for building learning games with assessment data, and for game designers not interested in going that far, how the Balanced Design framework can bring more coherence and deeper learning to your learning game design. But why else put in the effort with this approach?

As a field, we have built some impressive learning games. But designing these games on a more coherent framework, and documenting that for the learners and educators who will use your game represents not just a promising step forward, but evidence of true 'alignment' of learning, task and evidence data, that largely has not been present in the field of learning games to this point. This lack of 'alignment' is framed quite well by Robert Mislevy and colleagues, the lead developers of ECD:

“One can not simply construct ‘good tasks’ in isolation, however, and hope that someone down the line will figure out ‘how to score them.’ ”¹

In reality, this happens all the time in learning game design. Often, great focus is given to the teaching and learning the content of the game, but often little if any thought is given on how understanding how the student is advancing in understanding that content. That job is often left to the teacher—or designers take the position that it's not critical because they simply are not interested in giving the learner an exploratory, playful experience with the ideas. But more and more educators are turning to game-based learning, and looking to use these tools in their classroom. While we certainly believe this is one of the key tasks of a teacher — to observe, assess, and appropriately guide the next steps for the learner — we can not just hand them a learning game that for many educators feels like a 'black box' of a learning experience, which they didn't have a hand in creating, and expect them to figure it out. Assessment is just a term for the 'other side of the coin' of learning. The two are one. Learning can be defined by two interrelated questions that drive a never-ending cycle of improvement:

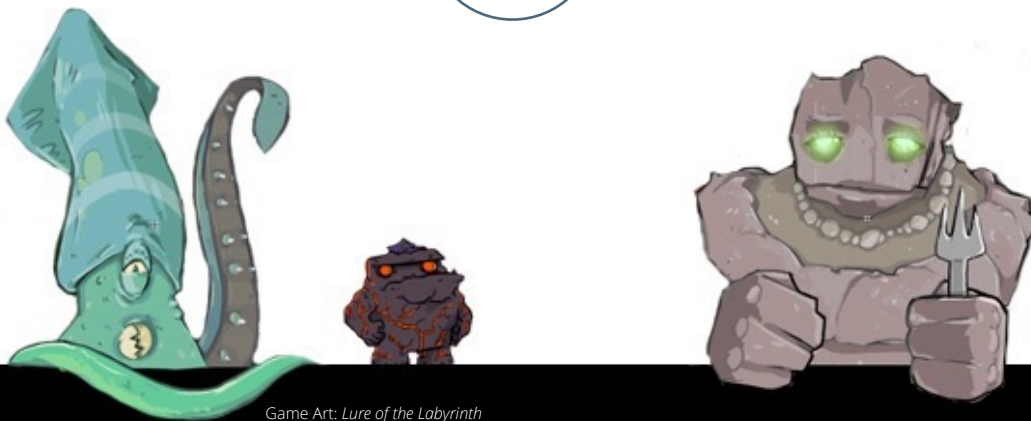
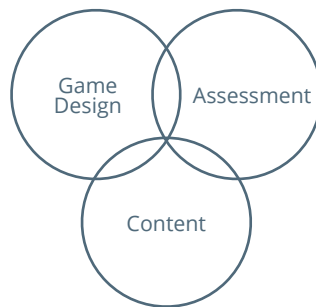
What are you trying to teach?

How do you know they know it?

Educators, and learners, need this feedback loop. As a field, we have largely ignored the second question. Balanced Design provides the structures to lay out this loop from the beginning of the design, and what is necessary in order to collect and display learning data based on game play for both educators and learners to use. This is quickly becoming one of the biggest goals — and barriers — for game-based learning in education.

But perhaps even more so, the second question is critical because it closes that feedback loop. But it is not just critical to the teacher or student because it helps them know where they stand, and what they might need next. Indeed, those are essential. It's critical because when we ask the second question at the beginning of the design process, it changes our foundation. Asking, and answering, that question, frames a content model and conceptual foundation for the design of the strategic learning experience in learning games. Without such an explicit and defined model, we are making judgments about the game design without a developed evidence model with the learning processes involved. As a result, the learning design comes from our best judgment rather than from evidence of learning.

For some learning game designers, the emphasis of the design and primary driver is the game's playful mechanics and design; whereas for others, the primary focus is on assessment and data collection. However, too often these approaches fall short in the other areas. Balanced Design argues for the coordination and alignment of these three areas — game design, content and assessment — for a more powerful learning experience and a more robust learning tool.





The playful learning Initiative

Helping educators use
games in the classroom.



“

Playful Learning is a national initiative to support educators in using games in the classroom. Funded by the Bill & Melinda Gates Foundation, the NewSchools Venture Fund, and powered by the Learning Games Network, the project aims to connect educators, and to support and empower them to connect with the resources and games that can have an impact in their classrooms. Over the past two years, we have worked with over 2,500 educators in small workshops and large 2-day summits across the US.

There are many educators who are game-based learning enthusiasts, who are already doing rock-star work with digital games in their classroom. But there's an even larger group of educators who are interested in game-based learning but aren't sure of what it's all about or how to get started. That's where Playful Learning can make a big difference.

One of the biggest complaints we hear from educators during Playful Learning workshops is the frustration with not being able to use games because the game doesn't give the learner, or the educator, any data on student learning and performance. Learning games are crafted learning experiences, and educators are master craftsman at creating learning experiences—that's why so many see the value in learning games. But as master craftsman, they are used to knowing what to look for in student behavior and performance or misconceptions in that experience. If they didn't create the learning experience (as is the case with learning games) and they don't have any data or feedback on what's going on, it's difficult for them to know how to proceed with the learner—and for many, difficult to justify the use of the game in class time.

“I would use games more in my classroom if they were able to give some kind of data on what the student is doing in the game or how they are progressing in their learning.”

7th grade teacher, Playful Learning workshop

In order to help connect educators with game-based learning resources, we began cataloguing many popular commercial as well as educational games, to map many of the critical details that educators would need or want to know in order to find or use a game—such as cost, platform, standards it's aligned to, etc. What was surprising to us was how many educational games either didn't have things like learning objectives or learning goals documented anywhere; or if they were, how each game did that differently and how difficult it could be to find. We realized how challenging this must be for educators—to find the right games for them, and to explore the support materials that vary so much across games.



Jennifer Groff
Project Director, *Playful Learning*



Peter Stidwill
Executive Producer, *Playful Learning*

”

A DESIGNER'S STORY

An XCD approach to game design



Jody Clarke-Midura, EdD, Assessment Lead

Louisa Rosenheck, Learning Game Designer

66

Since the earliest conception of Radix, it was important to us that the content and skills we were targeting be closely integrated with the game mechanics, or tasks, in the game. We also knew early on that we wanted to be able to provide teachers with detailed information on how their students were progressing and what content or types of problems they were struggling with.

We knew that using an Evidence-Centered Design method could be beneficial, and although we experimented with it in certain areas, we also knew that implementing the full ECD method requires a lot of time and resources—more than what we were able to put together for such a large game. For this reason, we instead used a simplified version of ECD, which didn't include such detailed data models but did guide our design process in useful ways.



Our first step was to decide which seven content areas in math and biology the game was going to target, and specify which concepts and skills we wanted players to explore. This was the Content Model. Once that was decided, we began designing the game environment and the tools players could use to interact with the world. From there, we were able to sketch out a narrative storyline and the quests that would make up the majority of the gameplay. These quests were comprised of tasks such as using tools, inspecting and collecting objects, creating new objects, and answering various types of questions.

A major part of designing these quest tasks was thinking about how they enabled players to explore the topic, and what evidence we could capture as they interacted with each task. This was our Task Model. To make it explicit, we went through the exercise of specifying which game actions engage players in each concept, which we called a 'learning objective.' To do this we made a list of content and practices such as "understand complete dominant traits" or "identify patterns in genetic traits" and matched each one with all the places in the game where we thought learning in those areas was happening. Thinking about what data could come out of those actions had a big influence on the design. For example we designed the trait examiner tool so that players have to initiate an action to find out a 'glumbug's' antenna length. This action could be recorded to provide information about what a student is doing in the game. In addition, when players breed a glumbug they can create and export a 'Punnett square object' that is stored in their inventory, can be examined by the student and also turned in for a quest objective. This means that when students submit the 'Punnett Square Object' as part of a quest, the game can record data on which parent bugs were bred and what trait the player was interested in.



While we designed the quest tasks, we focused on two things: what the learning objectives were and what kind of information or data we could observe as players interacted with those tasks. The more important connection to make was what this data could tell us about players' understanding of concepts and ability to use skills (i.e. how were the tasks providing information about whether the learning objectives were being met). This was our Evidence Model. If a player bred certain bugs could we be confident that they understood complete dominance? If they turned in an accurate Punnett square did it mean they were able to calculate the probability of the offspring's traits? All the data we could collect wasn't useful to teachers unless we could help them interpret it. Without conducting validity testing to correlate the actions with individual student knowledge, we were not able to say confidently what the game data meant about student competencies. However, exploring those connections throughout the design process enabled us to be clear about what the data could and could not tell us.

Aligning the content, tasks and evidence

After a first version of the tools and quests had been built, and once we had an idea of what knowledge and skills we wanted evidence for, it was time to go back and finalize the data collection piece. We did this by filling in a table that listed all the player actions we thought were significant in the game. For each action, such as turning in a quest or breeding organisms, we specified the inputs and outputs that we wanted recorded. That is, what decisions a player made before taking the action, and what the outcomes of that action were. Certain other data needed to be recorded along with every action, such as the player's ID number, location in the game, and a timestamp, so we laid that out as well. Collecting relevant data, even without a complex data model implemented on the back end, has provided teachers and researchers with important information for formative assessment of student progress.



The iterative process

Of course, as happens with any game, we encountered pitfalls during the development process. Some of the data we specified in our documentation was not able to be collected, resulting in an incomplete picture of player performance. In addition, in many cases we as designers were required to nail down the evidence requirements and data collection specs before the tools were far enough along to see how they really worked, or before we had a chance to playtest the final quests. This meant that we didn't always know to ask for all the data we would have wanted to collect. This points to a common flaw in the development process that many learning game designers know all too well, which boils down to the fact that *although we know how important testing and iteration is, there is rarely enough time to get it just right.*

Even without issues in development, designing for evidence in this way is never simple. In the design of Radix, striking a balance between open-ended activities and guided tasks proved challenging. We wanted to give players freedom to explore the world and make their own choices, but doing that meant that the game couldn't always know what the player was trying to do at a given moment. Logically, not knowing a player's goal makes it hard to interpret the data about what they were doing. On the other hand, while more constrained quests would provide more information on players' goals, it could mean that players had a more similar trajectory through the tasks, providing less individualized information about their understanding or misconceptions. This tension requires careful game design and assessment design in order to balance both goals. We found that the XCD framework provided a language and a method to weigh both sides in any given quest and explicitly address these issues.

Experimenting with the tension here made it clear that the appropriate balance for both engaging gameplay and evidence of learning lies somewhere in between those two extremes, and in Radix we feel that different quest lines achieved that balance to varying degrees of success. What did become clear throughout this process is that *the gameplay and assessment pieces of a learning game are so intertwined that it is nearly impossible to design or build either one before the other.* Rather, the connections between them and the target skills need to be made explicit from the beginning of the design process and revisited throughout it.



A DESIGNER'S STORY

The evolution of our game design approach



Mike Beall, Project Leader & 2D/3D Artist

36 We really stumbled our way into using this approach to game design. It began with a game called Progenitor X, and we started that game's design the way we would any other—starting to learn about the content, talking to subject matter experts at the university, etc. Progenitor X is about regenerative medicine, so we talked with some medical groups and two professors here to refine the learning goals and better understand the main components of stem cells. From talking with these experts and spending some time in the medical labs, we felt that the content really fit well with a puzzle game format. The learning goals should ideally reflect the concepts, so our thinking was that if we designed the game well and the player successfully completed the game it meant that they understood the concepts.

We moved forward with development as we usually would, early iteration with user testing, etc. Towards the end of the development, a researcher (and former educator) here at the Games+Learning+Society center came in to check out the game and started asking some really important questions, like, “how do you know the player learned X?” and “what does it mean when player does X versus Y?” ...the questions were logical, but we didn't have any way to really know with any certainty. We just started scratching the surface on what we could potentially do with the collected data, and it wasn't until the game was nearly done that we really went back and looked at these questions. We tweaked a few things at that point to collect a few pieces of data, but we weren't really able to make many substantial design changes at that point—but, it really showed us how important it was to think about this earlier in the design phase.

Once we were far enough with development of our data collection capabilities and no longer relied solely on pre/post tests, which I feel are problematic at best for getting at this kind of data, we realized that if we had a solid plan as to how we would prove a player knew X from doing Y in the game, this could help inform the design. We started to gain a better understanding of how to design certain types of game mechanics to work in combination with a well designed data collection plan.

This thinking really influenced how we approached our next game design project. We realized that by going back and looking more deeply at the game and the content, and being concerned with how we could really know what players learned, we looked at the content of the game differently than we did at the start—which would really affect the game's design. Going forward, we started with this work upfront, and had learning designers and people more fluent in assessment and data working with us from the beginning. And our subsequent games, like Crystals of Kaydor, were much better as we got better at this approach.





Our process is constantly evolving, partly because we now include assessment goals to support the learning goals in the game's design. And you're always going back to those assessment and learning goals as you flesh out the design. Those assessment tools really define what successful learning in the content domain looks like—you can't not acknowledge that.

Obviously, there are many methods of designing games for learning, but by using this ECD approach, it can make it easier to zero in on game mechanics that work better for learning and assessment goals upfront--and actually save time in aligning the game mechanics with learning goals. It's certainly possible to build a good learning game without using these tools, but we've found it to often be more difficult to show an educator exactly what the student learned, and how they were successful at it. It's one of the biggest sticking points we hear from educators—they know your game is about a certain topic, but it really doesn't unpack exactly what the game is covering, how each student performed in the game, and most importantly, what that means about what they understand about that topic.

In reality, learning goals and assessment goals are two sides of the same coin—it's tough to be successful with one without the other. The analogy I use with our development team is that of a football quarterback and a wide receiver. A quarterback's throwing mechanics are one thing, and in many ways are independent and the wide receiver doesn't really have to worry about that. But it's the interaction between the quarterback and wide receiver that really matters. The same thing goes for a learning game. You can successfully convey ideas and concepts in a game, but you have to be sure the student performance in the game is showing how they are actually getting it. If you're just looking at the quarterback you're not looking at the whole picture. They are two independent systems, but they work together. So at the end of the day did the wide receiver make the catch? It's easy to tell in football, but how do you really know in a learning game? Maybe the quarterback's throw wasn't accurate, but maybe the receiver wasn't able to catch it. You have to worry about both.



My Game Design Doc
Introduction
Target Audience
Learning Goals
Assessment Goals
Assessment Mapping
Game Play Concepts
Progression
Game Features
Data Collection
Narrative Synopsis
Main Characters
Interface
Initial Questions

Future directions...

Moving Towards the Next Generation of Learning Games

The field of learning games is standing on the horizon of many exciting things ahead. Assessment and data are key themes in this field, and for good reason—they provide us with powerful tools to create better learning experiences, and to help us better understand learning in game-based environments and beyond. Data is an abundant resource in games, if we know how to use it. We are moving toward an era where clickstream data in games offers us a rich sea of information to use to our advantage—and to the advantage of our learners. Cutting-edge tools like ADAGE — a platform to support the collection and interpretation of click-stream data in learning games (see Appendix A for more information) — are painting a picture of a future where we can build a bridge between the game-specific learning mechanics and actual click-stream data we are seeing in the game to better inform our game design models. Such a future is promising, as the playful and engaging environments support the learning, but also support meaningful data about the learner, their understanding, and performance in the game.

The field of learning games will be measured by how well we achieve the fundamental goals of creating learning games as powerful learning experiences to help learners break through ideas in bold new ways.

Refining that craft will be an ongoing process. Balanced Design is not a prescription or a template, but a design framework and one tool that can help in that process. However it is our hope that this will enable a larger discussion within the field on how such a framework can be enhanced and benefit the advancement of the field of learning games, our work, and most of all, the learners for which we design our games.




QUANDARY

Appendix A

ADAGE Beta

<http://adageapi.org>

Assessment Data Aggregator for Game Environments

There is a growing interest in investigating how games help learners learn. However, many researchers struggle to make sense of the myriad of simultaneous events happening when players engage in digital games or how these events might provide insight into the learning taking place. In response, ADAGE (Assessment Data Aggregator for Game Environments) aims to formalize the data captured when learners play with games. Creating a standard for this data offers the potential to both make it easier to collect and analyze the data, and allow researchers and developers to create a common set of tools and methodologies for working with ADAGE data.

ADAGE is an open-source data collection and analysis framework that transforms real-time click-stream (telemetry) data from games and other digital learning environments into formative evidence of learning. These data can be used to identify patterns in play within and across players (using data mining and learning analytic techniques) as well as statistical methods for testing hypotheses that compare play to content models.¹ ADAGE offers a robust framework for defining and specifying how data from games is gathered, stored, and made available to educational researchers, game developers, and classrooms. Overall, the ADAGE framework provides rich, method-agnostic data, scales easily, and is flexible enough to use across a wide range of games and platforms (e.g., Mac, Windows, iOS, Android).

ADAGE has three main components:

1. a **conceptual framework** for identifying salient data for learning (even across game genres);
2. an **open-source data specification and Unity client** (available on GitHub) for implementing this framework in-game; and
3. an **open-source API** to import and export ADAGE data.

Assessment Mechanics

ADAGE powers the collection of salient game-based learning data through a framework of *Assessment Mechanics*. Assessment Mechanics are ADAGE structures built into the game that allow for research on play and learning. Understanding game-based learning requires two levels of assessment mechanics: one to trace the paths players take through a game, and the other to access the player experience of game play.² Squire emphasizes that games as designed experiences provide endogenous engagement for the player through “roles, goals, and agency”.³ Thus, in learning games, there are two core types of designed mechanics: one set related to progression through the gameworld, as an engaging learning context; another may be designed as more direct measures of the content the game is trying to teach.⁴ Ideally, these also overlap; good educational games meld learning mechanisms with the core mechanics of the game, where gameplay itself is the only necessary assessment.⁵

1 Halverson, R. & Owen, V.E. (2014). Game-Based Assessment: An Integrated Model to Capture Evidence of Learning in Play. *International Journal of Learning Technology*, 9(2), 111–138.

2 Schell, J. (2008). *The art of game design: A book of lenses*. US: Taylor & Francis.

3 Squire, K. (2006). From content to context: Videogames as designed experience. *Educational Researcher*, 35(8), 19–29.

4 Gee, J. P. (2005). Learning by design: Good video games as learning machines. *E-Learning and Digital Media*, 2(1), 5–16.

Salen, K. & Zimmerman, E. (2004). *Rules of play: Game design fundamentals*. MIT Press.

Clarke-Midura, J., Mayrath, M., Dede, C. (2012). Thinking outside the bubble: Virtual Performance Assessments for Measuring Inquiry Learning. In M. Mayrath, J. ClarkeMidura, D. Robinson (Eds.) *Technology Based Assessment for 21st Century Skills: Theoretical and Practical Implications from Modern Research*. New York, NY: Springer-Verlag.

5 Gee, J. P. (2012). Big “G” Games. <http://www.jamespaulgee.com/node/63>

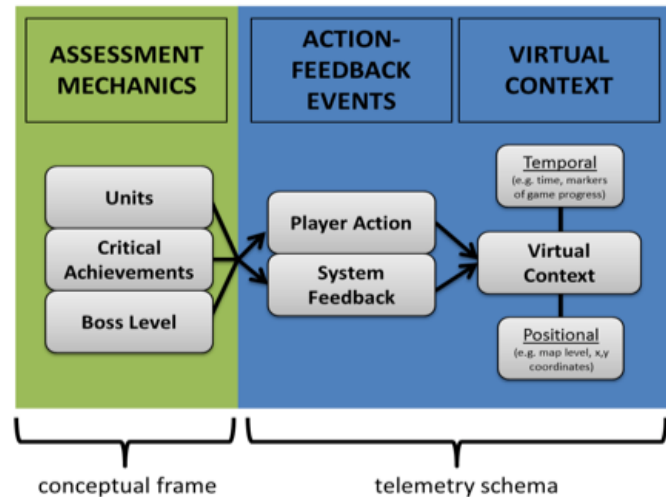
Shute, V. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J.D. Fletcher (Eds.), *Computer Games and Instruction*. Information Age: Charlotte, NC., 503–523.

The ADAGE framework identifies underlying game mechanics which serve as core occasions for player interaction. There are three base types of Assessment Mechanics:

- *Game Units* (capturing basic play progression)
- *Critical Achievements* (formative assessment)
- *Boss Level* (naturalistic summative assessment)

These Assessment Mechanics serve as data- collection (or assessment) anchor points, which yield data informed by core educational game design structures. This terminology also parallels concepts of formative and summative assessment in formal learning environments, and formalizes them as powerful elements of game design.¹

ADAGE constitutes the clickstream assessment (see figure on page 23); it captures the *assessment mechanics*, shapes the *telemetry schema*, and defines the learning telemetry and *final data output*. This telemetry parallels and supplements ECD task and evidence models with rich play context. This rich narrative and contextual data, which captures play progression often beyond the scope of a narrower task model, can be seen in the ADAGE diagram here.



Interaction with the designed assessment mechanics (left side) are manifested through player action and game event feedback cycles throughout the game (right side). These player actions and events are then seeded with rich context. For example, if a player finishes a quest (a unit) by engineering a brand-new healthy organ in the Progenitor X lab, the ADAGE system doesn't just record the end of a turn. To name just a few accompanying recorded features, ADAGE can also show what tool players used to build the organ, how quickly they built it, whether it was the correct organ to create, previously completed skill turns, previous success rates, overall game progress, and where in the gameworld the player was in at organ completion. This gives a context-rich data stream (supported by things like time, UI detail, and game location in the "virtual context" above) that can fuel ECD-based analyses *and* exploratory data mining. Conceptualization of these critical ADAGE framework pieces (assessment mechanics and context-rich telemetry) are folded into the embedded assessment design process shown above.

Through Assessment Mechanics, ADAGE operationalizes player interaction as the vital link between experience and game design (see green box in figure above). These three core AM types can easily overlap within a gameworld; they are not mutually exclusive, though they have distinct categories. Additionally, every game does not have to have all Assessment Mechanics in order to use ADAGE. Below we describe each mechanic, and connect it to ADAGE's underlying telemetry structure.

Game Units. The Game Units represent the core progress mechanic of the game. For example, in a game like World of Warcraft (WoW), the core unit is quests. By definition, game units have the property of being a repeating, consistent vehicle for making progress through the gameworld. Units can also be part of a hierarchy – for example, one set of quests may make up a particular map area, and completing all the maps means finishing the game. Thus, from broadest to smallest, Game Unit hierarchy might be: game-map-quest. The idea behind Units is that they are flexible enough to work across genres; Currently, the ADAGE Unit structure is applied to five different GLS games (*Progenitor X*, *Fair Play*, *Anatomy Pro Am*, *Tenacity*, and *Crystals*

1 Harlen, W. & James, M. (1997). Assessment and learning: differences and relationships between formative and summative assessment. *Assessment in Education*, 4(3), 365–379.
Gee, J.P. (2012). Big "G" Games. <http://www.jamespaulgee.com/node/63>

of *Kaydor*) and an interactive “game-like” museum exhibit (Oztoc¹) each with different genres and Unit types.

Critical Achievements (CA). In ADAGE, CAs are direct formative assessment slices of the content model. They are moments of direct content measurement within the context of normal gameplay. Seamlessly woven into the fabric of the game, CAs use naturalistic game mechanics to measure underlying educational content. For example, *Fair Play* is a GLS game that teaches about implicit bias in graduate education settings. In one *Fair Play* CA, the player needs to correctly identify a given bias to another character in order to progress. This is a direct demonstration of bias knowledge (as opposed to indirect movement through the learning context, like in game Units). The CA data structure aligns very well with ECD task analyses. CAs (analogous to a “task model” in ECD) are intended to be one kind of direct content assessment embedded in gameplay, looking at selected moments of performance as learning measures. Ultimately, CAs are a unique feature of educational games, and capture both learning AND play dynamics in the user experience.

Boss Level. The Boss Level is a final stage of a game that is a culmination of skills learned in gameplay. It is a naturalistic summative assessment, and can include both learning and progress mechanics (like CAs and Units). Gee notes that powerful embedded assessment occurs in “boss battles, which require players to integrate many of the separate skills they have picked up” throughout the game. Games are an ideal medium for this summative assessment, he asserts, since they can provide just-in-time performance feedback with low cost of failure. By formalizing the Boss Level as an Assessment Mechanic in ADAGE, we encourage deliberate inclusion of summative assessment in game design, and provide corresponding telemetry API structures for implementation.

Telemetry Framework

The Assessment Mechanics, informed by game design and assessment research, create a conceptual framework for identifying interaction data. The next ADAGE step moves us from concept (Assessment Mechanics, or AMs) to implementation (telemetry). The telemetry framework relies on the AMs to create a schema of context-rich data tags for implementation in the game code. Interpretation of student interaction often hinges on the context of the learning environment (in this case, the designed gameworld). The telemetry schema addresses this need by seeding the AM interaction data with vital contextual information.



Progenitor X is a zombie-themed puzzle game introducing stem cell regenerative medicine to middle school students, which originated and was built and researched at the Games+Learning+Society Center in Madison, Wisconsin in consultation with top scientists at the Wisconsin Institute for Discovery. The game was designed so that the core game design and narrative were aligned to the assessment mechanics—with the goal of developing a seamless, integrated, fully immersive play experience (i.e. a “designed experience”). *Progenitor X* is now one of the games that is collecting play and telemetry data to inform learning models on ADAGE.

www.gameslearningsociety.org/px_microsite

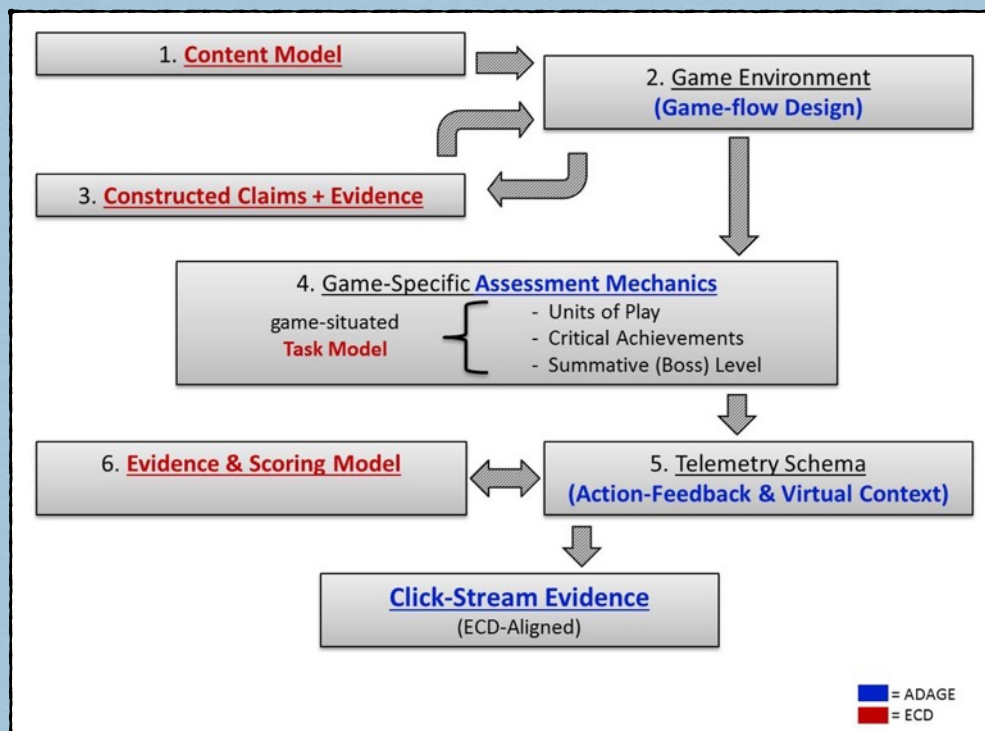


1 See *Oztoc Expedition* at http://www.gameslearningsociety.org/oztoc_microsite

The telemetry schema has two layers: an action-feedback layer and a Virtual Context layer (see figure on page 21). First, for each Assessment Mechanic, it identifies two sources of direct interaction: user action, and system feedback. It articulates the vital action-feedback loop that comprises of the interaction between the player and the game. The second layer, called the Virtual Context, attaches important contextual information to each action-feedback event. The Virtual Context can include important additional information like timestamp, map level, and screen x,y coordinates, adding a richer lens into the state of the game and the player when the event was recorded. These two layers work in tandem to provide researchers, teachers, and students context-rich telemetry data on Assessment Mechanics-based gameplay trajectories.

Feature Engineering & Analysis Lenses

ADAGE's context-rich data make ideal building blocks for feature engineering. Features are essentially variables of interest in the data, which can range from simple click locations to complex measures like accuracy over time. These features can then be used across a broad range of analysis techniques. Data lenses can include descriptive statistics, hypothesis-driven applied statistics, and machine learning techniques. Methodologies for hypothesis testing (like ECD) can use ADAGE data as dependent variables, independent variables, and covariates for use in associative or predictive modeling. Within a game design process, the figure below shows a clear mapping between ECD elements and ADAGE-supported assessment. The data produced by ADAGE also lends itself to data mining techniques often used with big learning data sets.



ECD-Informed Embedded Assessment Design in *Progenitor X*

ADAGE Tools to Support Implementation and Learning Analytics

To support teachers and researchers to integrate ADAGE into their learning game designs, we have developed several tools which can streamline implementation and provide insight into the learning taking place.¹ These tools have been bundled into three main “portals”: the ADAGE Researcher Portal, the ADAGE Analysis Dashboard, and the ADAGE Player Portal. The ADAGE Researcher Portal contains a suite of tools for collecting, querying, and analyzing data from ADAGE-enabled games. Through the research portal, researchers can quickly and easily create large cohorts of player accounts for use in playsquads, classrooms, or long-term studies, towards expediting the setup of large-scale data collection. The dashboard can also utilize QR code generation and recognition to allow players to easily log in to ADAGE on their Android or Apple devices. The data-querying interface allows researchers to construct basic queries within ADAGE, filtering and aggregating the data to create reports of player activities. These queries can be further saved for future use and modification. Data queries can be made at multiple granularities (e.g., on individual players or specific groups of players). The portal provides some basic data visualizations such as scatter plots, bar charts, line graphs and state diagrams, allowing researchers to easily and quickly explore the game data in different ways. These data can also be downloaded in a variety of formats such as CSV, JSON, and DTD (Datashop) for use in other third-party data analysis tools.

The ADAGE Analysis Dashboard allows researchers to upload ADAGE JSON files, transcript data, and researcher notes into a single page to support rich multi-modal analysis. Researchers can connect graphs of ADAGE telemetry data to timestamped transcripts and researcher notes to better triangulate their research data. On-screen annotations allow researcher to make connections between the various pieces of data to support their analysis. The ADAGE Analysis Dashboard will allow researchers to also perform Learning Analytics (e.g. develop Bayesian networks) on the telemetry, highlighting trends and insights in students’ learning data.

The ADAGE Player Portal contains a variety of information for players, teachers, and parents regarding player progress in ADAGE-enabled games. Using information supplied via the Researcher Portal, the Player Portal provides players with in-depth details about their in-game experiences through a base set of metrics which includes: personal progress and achievements, class and game leaderboards, and other relevant game statistics. With the help of researcher input, the player portal also provides detailed information on the correlation between player progress and the game’s learning goals.



Learning Games Playdata Consortium

<http://playdataconsortium.org>

ADAGE tools are being developed by the the Learning Games Play Data Consortium (PDC). Funded by the National Science Foundation (SMA-1338508), the PDC is a national consortium focused on building a research community and infrastructure for learning analytics in digital games for learning, connecting leaders in academia, learning games, and education.

ADAGE is open source with the following licensing agreement.

We welcome contributors and pull requests. Please join our developers group for more information.



**GAMES+
LEARNING+
SOCIETY**

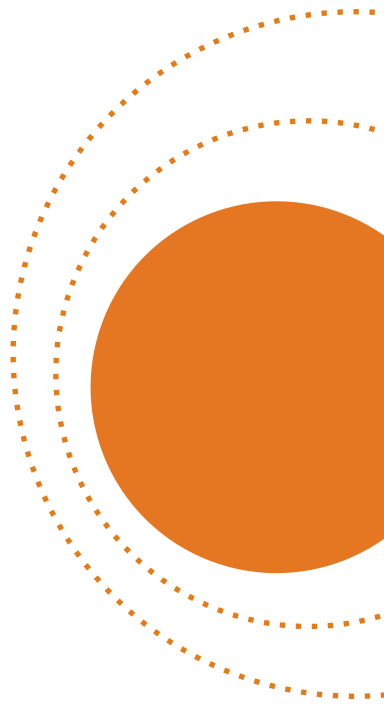


**LEARNING
games
NETWORK**

<http://adageapi.org>

© 2015 Games Learning Society

¹ Tissenbaum, M., Paiz-Ramirez, D., Stenerson, M., Goodin, I., & Berland, M. (2015). *ADAGE (Assessment Data Aggregator for Game Environments)*. Proceedings 11th Games+Learning+Society Conference. Madison, WI



Jennifer Groff, Ed.M.
Jody Clarke-Midura, Ed.D.
V. Elizabeth Owen, Ph.D.
Louisa Rosenheck, Ed.M.
Michael Beall

© 2015 Creative Commons License Attribution Share Alike 4.0 International,
Learning Games Network, MIT Education Arcade
Some rights reserved.
To view a copy of this license, visit www.creativecommons.org

LEARNING
games
NETWORK



Massachusetts
Institute of
Technology

Scheller Teacher Education Program
the education arcade