# Chapter 31
# Metrics in Simulations and Games for Learning

**Jan L. Plass, Bruce D. Homer, Charles K. Kinzer, Yoo Kyung Chang, Jonathan Frye, Walter Kaczetow, Katherine Isbister, and Ken Perlin**

## *Take Away Points:*

In this chapter, we will
1. Review the literature on embedded assessment in games for learning,
2. Introduce an approach to improving the diagnostic power of game metrics through learning mechanics and assessment mechanics,
3. Present a case study employing Cognitive Task Analysis for the study of learner behavior in science simulations,
4. Present two case studies illustrating the approach in learning games developed by G4LI.

## 31.1 Introduction

This chapter introduces the approach taken by the *Games for Learning Institute* (G4LI) to assess learning and related learner variables, with a focus on the use of metrics obtained during game play and simulation exploration. Learning is fundamental to all games (Gee 2008). At minimum, players must learn the basics of a game's mechanics to play. Additionally, players must uncover what these

J.L. Plass, Ph.D. (✉) • J. Frye • K. Isbister, Ph.D. • K. Perlin, Ph.D.
Games for Learning Institute (G4LI), New York University,
New York, NY, USA
e-mail: jan.plass@nyu.edu; jplass@gmail.com

B.D. Homer, Ph.D. • W. Kaczetow
Games for Learning Institute (G4LI), CUNY Graduate Center, New York, NY, USA

C.K. Kinzer, Ph.D. • Y.K. Chang, Ph.D.
Games for Learning Institute (G4LI), Teachers College Columbia University,
New York, NY, USA

M. Seif El-Nasr et al. (eds.), *Game Analytics: Maximizing the Value of Player Data*,
DOI 10.1007/978-1-4471-4769-5_31, 

mechanics are for, and what the game designer wants them to do (Cook 2006). Feedback mechanisms are an example of how game designers encourage (reward) or discourage (punish) a behavior. Game mechanics for learning must incorporate all of these aspects, from the moment-to-moment activities in which players engage, to reward and punishment systems.

In our research and game design work over the past several years we have found it helpful to use the following terms in the design of game mechanics (Plass et al. 2011):

*Learning Mechanics* describe essential high-level activities, grounded in learning sciences, that have learning as the primary objective;
*Assessment Mechanics* describe essential high-level activities, grounded in test theory, that have assessment as the primary objective;
*Game Mechanics* define the essential game play activity and can be based on learning mechanics, assessment mechanics, or both.

In our definitions, learning and assessment mechanics are meta-mechanics, i.e., descriptions of activities that become the foundation for corresponding game mechanics, following criteria to preserve their intended learning or assessment objective. That is, learning mechanics and assessment mechanics are constraints communicated to game designers who then select or design one or more corresponding game mechanics that accommodate these constraints.

Even though they do not use our terminology, commercial games make frequent use of our concepts of learning and assessment mechanics, though not to the extent required for games for learning. Learning mechanics are often used in commercial games for the tutorials rather than for conveying educational content. Let's use the example of *Mirror's Edge*, a first-person game that includes many complex maneuvers to traverse across rooftops and other urban terrain. The tutorial utilizes learning mechanics of *modeling,* which are grounded in the theory of Cognitive Apprenticeship (Collins 1988; Liu 1998). In addition to telling players what buttons to push, the game has a non-player character (NPC) model the action that needs to be performed. There are several game mechanics that are required to make this learning mechanic work, such as creating the NPC model's AI and positioning the camera for proper observation. All these game mechanics, as well as others, are implementations of a single learning mechanic.

Assessment mechanics in commercial games are often used in tutorials and throughout the game to ensure that players have learnt a certain feature and are using it correctly. In these kind of cases, game metrics are helpful in the design and playtesting process to assure that the experience is well designed, is at the correct difficulty, and is enjoyable overall. Other games, such as *America's Army* (AA) use metrics to rate players across several "army values"; including loyalty, duty, respect, selfless service, honor, integrity, and personal courage (America's Army 2012). Since these values are complex, assessment mechanics describe the player activities that can be used to assess the level of each of these values for any given player. These activities are then used by game designers to design the AA game mechanics. In the case of the complex AA values, each assessment mechanic was implemented through a combination of several game mechanics. For example, for a player to earn points in loyalty, they must perform in-game actions such as mission objectives or neutralizing enemies while being connected to a teammate.

The purpose of using assessment mechanics in games for learning is to improve the diagnostic power of metrics collected during game play. Assessment mechanics help game designers select or design game mechanics that generate useful game metrics that allow us to measure variables related to learning, including learning outcomes (cognitive, behavioral, social, affective), trait variables, general state variables, and situation-specific state variables. In our approach, which we describe in this chapter, log data of game events and user behavior are also supplemented with observational data (video recordings or observer protocols) or eye gaze data that allow for triangulation of findings and therefore create more valid assessments of these variables of interest.

We believe there are two main benefits to our approach. First, embedded assessment, which uses measures based on metrics collected during game play, enables us to gain more detailed insights into learning than many traditional instruments allow with respect to the processes of learning and learning outcomes. This has implications for research as well as learner competency testing. Second, by using assessment mechanics to measure a series of learner variables, a learner model can be developed that allows us to design games that are individualized and adaptive to a learner's specific needs and characteristics. This has implications for the design of effective games for learning. In this chapter, we will discuss these implications and describe which variables to log, and which learner variables can be measured using game metrics. We will also provide illustrative case examples.

While most of this chapter concerns learning games and simulations, considerable progress has been made in using game metrics and telemetry in commercial games, examples of which have been discussed in this book. More examples include the Tracking Real-Time User Experience (TRUE) method (Kim et al. 2008) developed by Microsoft Game Studios, which combines several HCI approaches to track user data in a comprehensive system. The system tracks user actions and stores them in time-stamped log files that allow data to be analyzed as a sequence of events and "event sets", which adds contextual information that helps make sense of the event (Kim et al. 2008). Brief surveys are instrumented into games to collect attitudinal data, which allows direct user feedback. Data visualization within the TRUE system involves the creation of a meaningful hierarchical scheme and presenting the data that allows for navigating from high-level schemes to details and also linking between sources of information, such as log file data or video recordings of players (Kim et al. 2008).

Research has also looked at defining patterns of play using game metrics to identify player personas, which provide insight into the play-styles players consistently exhibit (Tychsen and Canossa 2008). Other research on commercial games has examined the use of time-spent reports (DeRosa 2007) and patterns of social behavior (Ducheneaut et al. 2004). Valve, the development studio behind *Half Life*, has also presented on the use of game metrics, survey data, and biometrics, such a skin conductance and eye tracking to gain a better understanding of players (Ambinder 2009). In addition, the many chapters of this book present several case studies of the use of

**Table 31.1** Examples of variables that can be measured using game telemetry

| General trait variables | General state variables | Situation-specific state variables |
|---|---|---|
| Executive functions | Achievement goal orientation | Emotional state |
| Spatial ability | Self-regulation, metacognition | Cognitive load |
| Verbal ability | Prior knowledge in the specific subject matter | Engagement |
| | Attitudes and interests | Situational interest |

metrics (see Chaps. 4 and 8), their types and definitions (see Chap. 13), their visualizations (see Chaps. 18 and 19) as well as triangulation techniques (see Chaps. 21 and 22). In contrast, the work we are presenting in this chapter investigates variables and methods similar to those used by commercial game designers, but views usability, fun and engagement through the lens of learning game designers.

This chapter is intended for game designers interested in the design for games for learning, ideally with some background in the learning sciences. There are many questions related to game telemetry in practical applications that are beyond this chapter's scope, such as how and when to introduce assessment tools in production, how to create a practical user interface to present the data, and how to prioritize assessment vs. content development given the constraints of finite resources.[1] Some of these issues, such as data reporting and visualization, are covered in Part IV of this book, other topics strictly pertaining to the educational domain will have to be addressed in other publications.

## 31.2 Review of Embedded Assessment in Games for Learning

The first two questions that must be considered when planning any assessment are: *what* is being assessed, and *why* is it being assessed? In other words, what variables do we want to measure, and what is the purpose of collecting this information? For educational materials, learning outcomes are the most common variables assessed to determine the effectiveness of an educational intervention. Learning outcomes are a way of conceptualizing the knowledge, skills, and abilities that the learner gains by playing the game. However, there are many other variables that may also be of interest. We have identified three broad categories of educationally relevant assessment variables: *General Trait Variables*, *General State Variables*, and *Situational-Specific Variables*, see Table 31.1.

Traits are relatively stable and are typically not targeted for change in educational games, but are often of interest and can be affected by game play (Anderson and Bavelier 2011). These include variables such as learners' abilities (spatial, verbal, etc.)

[1] We thank Bill Shribman, one of our very thoughtful reviewers, for raising these questions.

and capabilities (e.g., executive functions). State variables, such as prior knowledge of subject matter, attitudes, self-regulation and meta-cognition, and goal orientation, are more malleable. They are therefore often the target of change in education games. Finally, situational variables are more transient and are often the result of learners' interactions with the game. They include emotional state, cognitive load, situational interest, and engagement. Effective educational materials, including learning games, are designed to influence situational variables in ways that maximize learning outcomes, for example, by reducing extraneous cognitive load and increasing engagement. By assessing these variables, one can determine if design choices have had the desired effect on relevant situational variables.

The purpose of assessment can be divided into two broad goals: description and evaluation:

*Descriptive assessment* provides a picture of the learners using an educational game. In general, descriptive assessment measures trait and demographic variables (such as age, gender, etc.) through questionnaires or surveys provided prior to engaging with educational materials.

*Evaluative assessment* attempts to measure something that is directly relevant to – and influenced by – the learning process. Evaluative assessment can be further broken into two types:

- *Formative assessment*, part of the instructional process; provides critical information to inform the learning process, and
- *Summative assessment*, typically conducted at the end of an educational intervention to determine if the intervention was successful. By combining descriptive and evaluative assessments, it is possible to determine if design choices are working for all learners.

Once it has been decided what will be measured and why, one must determine *how* to measure the variables of interest. In learning games, assessment can be embedded within the game's environment, making it unnecessary to employ survey and other paper-and-pencil measures for variables that can be measured based on the data collected by the game in user logs. Of course, there will always be variables requiring subjective observation or testing outside the game environment, perhaps to determine whether learners generalized the tasks taught by the game, or whether they can transfer knowledge to solve out-of-game problems. In the following section, we will discuss the ways in which this *embedded assessment* can be developed, and review some examples from the research literature.

### 31.2.1 Evidence Centered Design: Layers and Models

The most relied upon framework for developing embedded assessments in educational games is *Evidence Centered Design* (*ECD*; Mislevy et al. 1999, 2002, 2003). ECD requires designers to be specific about what skills, knowledge, or other traits are being assessed, and what learners need to do, say, or create to provide evidence relating to the variables being assessed. Although ECD can be applied to many assessment scenarios, it is particularly useful for guiding performance-based assessments, such as embedded assessments in learning games.

There are several layers of assessment design decisions in the ECD framework, including:

- *Domain Analysis* layer, where information is gathered about the domain of interest.
- *Domain Modeling* layer, where narrative assessment arguments are built, based on information gathered during domain analysis.
- *Conceptual Assessment Framework* layer, where narrative assessment arguments are translated into guidelines for specific tasks.
- *Assessment Implementation* layer, where assessment related tasks are presented to learners and responses are analyzed.
- *Assessment Delivery* layer, where learners' interactions are coordinated with assessment tasks and assessments are scored and reported.

### *31.2.2 ECD Assessment: A Closer Look*

ECD assessment layers are guided by the *conceptual assessment framework*, consisting of three models in which relevant activities take place. The first is *Student Models*, which identifies relevant *skills, knowledge, identity, values, and epistemology* within a specific learning domain (Shaffer 2006). The student model ranges from simple, even including only one variable, to complex, involving multivariate item response theory or latent class modeling (Mislevy and Riconscente 2005). *Evidence Models* describe what learners must say, do, create, etc. to provide evidence related to the skills, knowledge, etc. being assessed. They also provide scoring rules and statistical models for how observed data relate to components of the student model. *Task Models* specify how variables from the evidence model will be collected in the educational game. What will learners say, do or create to provide evidence of the skills, knowledge, etc., being assessed? ECD also includes a *presentation model* that describes how relevant materials are presented to learners throughout the assessment process. Although the different ECD models can be considered separately, their interconnectedness means that effective assessment requires them to be considered jointly and revised iteratively (Rupp et al. 2010).

A number of the assessment approaches used in educational games are either derived directly from an ECD approach, or fit easily into the ECD framework. For example, Shute and her colleagues (Shute et al. 2009) used ECD to seamlessly embed assessments into educational games in an approach they call *stealth assessment*, because the assessment is not obvious to the learner.

With stealth assessment, Bayesian networks are used in the evidence model to determine how performance in the educational game relates to specific variables in

**Model–Based Assessment: Examples**

Shute and her colleagues have conducted a number of studies using ECD and Bayesian networks to embed a stealth assessment into commercial and education video games. Shute et al. (2009) assessed creative problem solving in *Oblivion*, a commercial roleplaying game where players encounter a river they need to cross that was full of dangerous fish. The ways that players solved this river-crossing task was evaluated to assess players' creative problem solving skills. Two *Oblivion* experts rated possible solutions with respect to *novelty* and *efficiency*, two concepts identified as being important for creative problem solving.

The concept model of creative problem solving identified two main sub components, problem solving and creativity (Shute et al. 2009). Both have elements of efficiency, which Shute et al. defined as the quality and quantity of the steps taken to reach a solution. Novelty, defined as choosing low frequency solutions, was considered independent of problem solving. The evidence model for this assessment had to define the connections between observable behaviors, creativity and problem solving. The action model defined interactions between the student and the game that were used to glean data. Each action taken to solve a problem (recorded in log file metrics) was scored on novelty and efficiency. Bayesian networks were used to covertly assess the river-crossing task. This approach has the potential to analyze tasks in games that allow experts to define the concept model and devise scoring rules (Shute et al. 2009; Shute 2011), with the possible exception of emergent gameplay where all possible solutions cannot be known initially and evaluated.

the competency model (Shute and Kim 2011). Bayesian Networks are directed graphs with probabilities attached to each node (Mislevy and Gitomer 1996). Each node corresponds to a variable of interest from the student model. The first step in implementing a Bayesian network is to define the structure of the relevant content domain (Mislevy and Gitomer 1996). Next the structural relationships must be transformed from relationships between constructs to relationships between probability distributions (as part of the conceptual assessment framework). These probability distributions are computed using empirical evidence and/or theory. Distributions of variables can then be obtained based on other variables that are at a level above the variable of interest. The probability of a given event is calculated recursively, as data is fed into the system (Mislevy and Gitomer 1996). A dynamic Bayesian network functions in a similar way, but also considers time as a variable (Iseli et al. 2010).

While the use of Bayesian networks is part of ECD's conceptual assessment framework, *Cognitive Task Analysis* (Schraagen et al. 2000; Williamson et al. 2004) is used primarily to support the domain modeling layer. CTA identifies the knowledge representation and cognitive processes that underlie complex behaviors, such as job performance or problem solving (Williamson et al. 2004).

Although the specific details of how CTA is carried out can vary, it generally begins with the identification of the tasks required to perform a target action or achieve a learning goal. The skills and knowledge necessary to accomplish these tasks are then identified, most often by developing a visual representation (Lajoie et al. 1998; Roth and Woods 1989). Once the tasks and the underlying knowledge representation are identified, a variety of techniques are used to have learners verbalize their cognitive and metacognitive processes as they perform the tasks (Cordingley 1989; Roth and Woods 1989). The information provided during the vocalizations is then used to inform the design of assessment tools. While many CTAs are completed manually, there are approaches to automate them (Shute and Torreano 2003). This ECD approach has been used to create embedded assessment in a number of educational games. However, there are other approaches that did not explicitly use the ECD framework, such as described by Heeter et al. (2009).

## 31.3 Case Study: Assessment of Meta-cognition in Simulations Using User Logs

Although Bayesian networks are particularly useful when the variables of interest are complex and their behavioral expressions are not well defined, Bayesian networks are not always needed, particularly for variables with clearly identified behaviors. This is the situation with our first case study, which utilized the *Behavioral Measure of Metacognitive Processes* (BMMP), a method to assess learners' metacognitive processes from metrics data. We developed this approach, a modified version of CTA, to examine students' metacognition as they interacted with science simulations (Chang et al. 2008; Chang and Plass 2012). An advantage of conducting assessment with metrics data is that it can provide causal explanations of why and how learning was effective (Winne 1982) by capturing the process through which learners demonstrate their development of specific competencies (Nelson et al. 2010).

> *Metacognition* comprises higher order thinking, reasoning, and learning skills (McCombs 2001); this process is employed by learners to direct their learning processes. Metacognition affects learning by influencing how students use resources in the learning environment (Aleven and Koedinger 2000; Hill and Hannafin 2001), and by mediating the appropriate use and regulation of cognitive resources (Schraw and Moshman 1995).

### *31.3.1 Description of the BMMP*

The BMMP is a framework for the construction of a rubric to identify metacognitive processes from metrics data, i.e., from user behaviors captured in log files (Chang et al. 2008; Chang and Plass 2012). The resulting BMMP rubric is a set of behavioral patterns that manifest metacognitive control within the specific learning environment. The BMMP can be applied to different student-directed exploratory learning environments, including games, simulations, and microworlds.

The large quantity of metrics data poses challenges to data analysis. By employing Cognitive Tasks Analysis (CTA), the BMMP can address the limitations associated with using metrics data, namely the difficulty of observing internal processes through manifest forms, and the selection bias in the variables being observed (Efklides 2002; Perry et al. 2002). CTA is able to identify the cognitive structures and processes beyond the observable behavior (Brown 1997), provide rich descriptions of implicit psychological processes and reduce selection bias by ensuring that all relevant cognitive strategies are included.

We established the validity of the BMMP in a study in which BMMP data was triangulated with think-aloud data. In addition, our research has shown that the BMMP was able to reliably assess learners' metacognitive processes across different learning environments (Chang et al. 2008; Chang and Plass 2012).

BMMP begins by identifying the learning goals addressed in the learning environment. Then, a CTA is conducted to develop a rubric, identifying the behavioral evidence of metacognitive control by determining: (1) the tasks required to accomplish the learning goal, (2) the skills and knowledge necessary to accomplish these tasks, and (3) the observation of participants on task, often accompanied by interviews, concurrent verbalization, or verbal prompts. The BMMP adopts the CTA's first two phases to identify the tasks and metacognitive strategies required to accomplish the learning goal. The result is a set of behavioral patterns, or a coding rubric, which can identify metacognitive processes from the metrics data.

#### 31.3.1.1 Case Example–Assessment of Metacognitive Processes with Interactive Computer Simulations

We applied the BMMP to assess high school students' ($N$= 457) metacognitive processes while learning with two interactive computer simulations for chemistry, one on Kinetic Molecular Theory and the other on the Ideal Gas Laws. The relationship between students' prior knowledge, metacognitive processes, and the learning outcome, especially in comparison to the self-report measure of students' regulation of their learning processes, were examined. Both simulations involve a set of variables related to molecular movement. Understanding the relationships between the variables is essential for understanding the chemistry concepts. With the simulations, students can explore the chemistry phenomena by: (1) setting the hypothesis about the relationship between the variables, (2) experimenting with the variables by

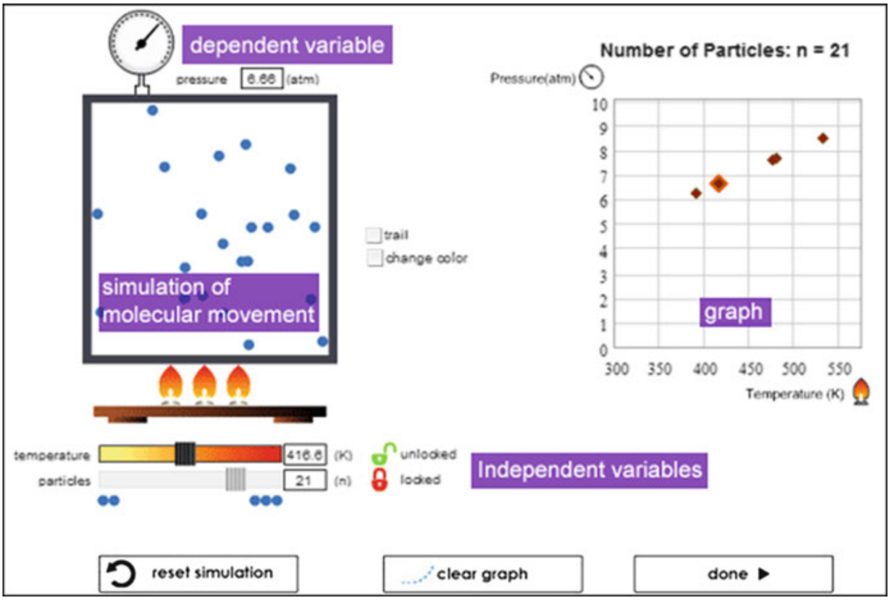


**Fig. 31.1** Interactive computer simulation for kinetic molecular theory

directly changing the value of independent variables, and (3) observing the results through a visual simulation of molecular movement and a graph that displays the results of the experiment (Fig. 31.1).

We assessed students' *prior knowledge* using a knowledge pre-test, and *Self-Regulation* using a self-report measure, the *Motivated Strategies for Learning Questionnaire* (MSLQ; Pintrich et al. 1993). The BMMP assessed the quality of metacognitive control based on the metrics data from simulations. Finally, learning outcomes were measured for comprehension and transfer of knowledge.

## Development of the BMMP Rubric

*Step 1: Analysis of the Learning Goals.* First, a content analysis for each of the interactive computer simulations was conducted to identify learning goals. Both simulations' learning goals consist of understanding the relationships between independent variables (e.g., temperature) and dependent variables (e.g., internal pressure, rate of diffusion) (see Table 31.2).

*Step 2: Cognitive Task Analysis (CTA).* A Cognitive Task Analysis identified the behavioral representation of metacognitive control and the conditions under which such patterns can lead to the achievement of the learning goals. The result of the CTA was a rubric containing behavioral patterns that manifest metacognitive control in a given learning environment.

**Table 31.2** Learning Goals for M&M Simulations

| Simulation | Learning goals |
|---|---|
| Ideal gas laws | Internal pressure is directly proportional to temperature<br>Internal pressure is inversely proportional to volume<br>Volume is directly proportional to temperature |
| Kinetic molecular theory | Internal pressure is directly proportional to temperature<br>Internal pressure is directly proportional to number of particles<br>Matter (gases) are made up of particles in constant random motion |

**Table 31.3** BMMP rubric for the first learning goal of Ideal gas laws simulation

| Behavior pattern | Condition | Strategy | Observation |
|---|---|---|---|
| Temperature (T) – T … | Two or more consecutive manipulations | Experimentation and observation of relationship between T and P while Volume (V) is locked (held constant) | T increase, P increase<br>T decrease, P decrease |
| Pressure (P) – P … | Two or more consecutive manipulations | Experimentation and observation of relationship between P and T | P increase, T increase<br>P decrease, T decrease |

With our simulations, the learning goals consist of relationships among pairs of variables (i.e., one independent and one dependent variable). The first learning goal of the Ideal Gas Laws simulation was to understand that *internal pressure is directly proportional to temperature*. At least two consecutive observations are needed to identify a rule defining the relationship between each pair of variables (Gick and Holyoak 1983; Schwartz and Black 1990). Therefore, student behavior of manipulating the temperature variable two or more times consecutively while holding the volume variable constant was recognized as reflecting a metacognitive strategy for this learning goal (see Table 31.3).

*Application of the BMMP Rubric.* The resulting BMMP rubric was applied to code metrics data and compute a BMMP measure. BMMP measures for the simulations were computed as a sum of two proportions to reflect the two types of metacognitive control required to accomplish the learning goals of the simulations: *adaptive behaviors*, i.e., the number of adaptive behaviors as identified by the BMMP rubric over total behaviors, and the *number of adaptively explored variables* over all variables available for exploration.

For both simulations, prior knowledge and metacognitive processes evaluated via the BMMP predicted learning outcomes. Interestingly, the self-report measure of self-regulation did not predict learning, suggesting that the BMMP-based behavioral measure is better able to capture learning relevant metacognitive processes than a paper-based survey of self-reports; see Chang (2010) for details.

This case example shows that CTA is a feasible way to establish a task model for measuring metacognition. The resulting BMMP measure was able to predict learning in two different simulations, whereas a self-report measure of metacognition was not. In comparison to open-ended sandbox games, the simulations used for this study are relatively limited in their exploration options. It is therefore reasonable to expect that an application of this approach to an environment that provides more opportunities for self-regulation will yield a highly valid and reliable way of measuring learners' metacognition using game metrics.

## 31.4 G4LI Approach to Assessment in Games for Learning

In this section, we will discuss the approach for assessment of learning and related learner variables taken by the Games for Learning Institute (G4LI). We will first define game mechanics in general, and then introduce the concepts of learning mechanics and assessment mechanics. We next describe how our approach extends the ECD model by adding observational data to the task model. We also provide case examples of how this learning and assessment approach was applied in two games, and summarize the studies we conducted with these games to identify initial design patterns for games for learning.

### *31.4.1 Game Mechanics*

The concept of *game mechanics* is central to understanding games. Game studies scholars and developers have offered definitions that range from finely detailed to broad and conceptual approaches. In general, game mechanics are "the various actions, behaviors and control mechanisms afforded to the player within a game context" (Hunicke et al. 2004, p. 3). In a narrower sense, a game's core mechanic is "the essential play activity players perform again and again and again" (Salen and Zimmerman 2003, p. 316). In other words, a game's core mechanic contains the moment-to-moment actions and interactions in which the player engages while playing the game. Finally, a definition proposed by Cook (2006) describes game mechanics as "rule based system/simulations that facilitate and encourage a user to explore and learn the properties of their possibility space through the use of feedback mechanisms" (p. 1). Game mechanics are, therefore, a means to guide players into particular behaviors by constraining the space of possible plans to attain goals (Järvinen 2008). Game mechanics do not only apply to players, as they are also defined as "methods invoked by agents, designed for interaction with the game state" (Sicart 2008), a definition that includes the game AI as actor, in addition to the player.

### *31.4.2 Learning Mechanics*

When games are designed with the explicit goal of facilitating learning, game mechanics must go beyond making a game fun and engaging–they must engage players in meaningful learning activities. The game mechanic becomes an integral part of the learning activity. Game designers have long seen this connection and have argued that new mechanics are needed to engage the player in a way that facilitates learning (Isbister et al. 2010). Most importantly, the designers interviewed by Isbister et al. made a strong case that learning needs to be embedded in a game's core mechanics rather than added on to existing mechanics. Game play cannot be used as a reward for answering questions; and vice versa, questions cannot be forced into unrelated game play.

> An example of poor integration of learning into a game is when a learning game uses an established game mechanic, such as a shooter mechanic, and the learning mechanic consists of a popup question that must be answered before players can resume the game.

To emphasize this qualitative difference of mechanics that are designed for learning, we offer the following definition of what we call *learning mechanics*:

> Learning mechanics are patterns of behavior or building blocks of learner interactivity, which may be a single action or a set of interrelated actions that form the essential learning activity that is repeated throughout a game.

Learning mechanics adapt the moment-to-moment activity of a game mechanic into a meaningful learning activity, often expressed as player choice. All games offer players series of choices and then react to those choices with new challenges. Learning mechanics can push game mechanics further to offer players choices that help them learn as well as facilitate gameplay. A game's learning aspects become an integral part of the game play rather than an addendum to the game mechanic.

The relationship between game mechanics and learning mechanics is that learning mechanics are meta-mechanics that become the foundation for the design of a corresponding game mechanic or collection of game mechanics. Learning mechanics are not themselves playable mechanics. They describe the functions of the actions available to players in the game, but they don't describe the actions themselves. For example, the learning mechanic might specify that the learner/player should be able to apply rules to solve problems, but does not describe the corresponding game mechanic. Concretely speaking, in the game Angry Birds there are three core gameplay mechanics:

– the slingshot system,
– the simulated laws of gravity affecting birds, pigs and other pieces,
– the physical static simulation applied to all non player-controlled elements consisting of mass, structural solidity and friction.

The learning mechanic in this case could be based on a constructivist approach, or simply a "learn by doing" principle:

– match properties of different objects (here, different types of birds: black/bomb, yellow/acceleration, green/Boomerang, etc.) to properties of the different materials (here, materials of structures protecting the pigs: glass, wood, stone),
– explore the rules that bind trajectory, velocity and mass.

In this *Angry Birds* example, these two learning mechanics are instantiated through the combination of the three core mechanics described above. In the context of another game, however, the same learning mechanics could be instantiated by using other game elements and mechanics, such as different types of air planes for which players draw the flight path, such as in *Flight Control*, or using jet packs to guide your character to a specific place, such as in *Little Big Planet*.

Learning mechanics are meta-mechanics that can be formulated independent of a specific game (though often linked to a specific game genre), and that need to be instantiated as game mechanics to describe the concrete actions and their affordances for the players in the system that the game represents. Playing the game equals learning these tools and moves, becoming familiar with them, and having the satisfaction of solving challenges, of "beating" the game (Juul 2003). Additionally, in designing game mechanics based in learning mechanics, game designers consider the game feel, the feel of engaging the core mechanics through interactive elements, visual elements, emotional elements and sound elements (Swink 2008).

#### 31.4.2.1 Criteria for Effective Learning Mechanics

Several criteria can ensure that Learning Mechanics will be effective in facilitating learning. Most importantly, learning mechanics are grounded in learning sciences and learning theory; many such theories and frameworks have been developed that can be used as the basis for the design of learning mechanics. Examples include *Cognitive Flexibility Theory* (Spiro et al. 1988; Spiro and Jehng 1990), *Cognitive Apprenticeship* (Collins 1988; Liu 1998), *Anchored instruction* (CTGV 1990, 1993), and *Situated Learning* (Lave and Wenger 1990). From these and other, related theories, designers choose activities that engage the learner in meaningful interaction with a specific subject.

The interactions in learning mechanics should be designed from a clear model of interactivity types. For example, the INTERACT model distinguishes three types of interactivity: behavioral, cognitive, and emotional interactivity, and describes their relation in learning processes, such as feedback and guidance (Domagk et al. 2010). Learning designers use these interactions to describe actions that allow learners to generate solutions to the problems that are designed to facilitate learning. Models like INTERACT help designers identify what behaviors (e.g., clicks, controller button actions, or touch actions) and emotion to engage the learner to facilitate thinking and decision making. This is a noteworthy difference to definitions of game mechanics, which question a deterministic relation between inputs, controls and mechanics (Järvinen 2008; Sicart 2008).

#### 31.4.2.2 Requirements for Designing Game Mechanics Based on Learning Mechanics

In the creative process of instantiating learning mechanics as game mechanics, game designers must ensure that the learning goal is preserved. Thus, when selecting

**Table 31.4** Library of learning mechanics and associated game mechanics

| Learning mechanic | Corresponding game mechanics |
|---|---|
| Reciprocal teaching: learner teaches target concepts to NPCs | G4LI *Noobs v. Leets* reciprocal teaching |
| Learner places icons representing key concepts to solve problems | G4LI *Supertransformations*! Problem solving using reflection and rotation icons |
| Learner creates authentic problems to solve by other players | G4LI AR Simulation Game for Science Learning mechanic of creating locations where specific scientific principles were applied |

a game mechanic to implement a particular learning mechanic, designers need to consider several requirements, including:

1. *Don't introduce excessive demands by unrelated information that distract from the learning goal*. For example, when the goal is to solve a science problem, the game narrative should not present excessive amounts of unrelated information.
2. *Don't introduce excessive demands by unrelated tasks that distract from the learning goal*. For example, when the goal is to collect essential information in the game, the collection task should not be so difficult that the learner cannot move on.
3. *Keep it challenging–don't excessively reduce the amount of the required mental effort to process the game's essential learning content*. For example, when solving algebraic equations, the game should not automatically indicate the changed sign of a term that is being subtracted.

A more detailed discussion of these requirements is beyond the scope of this chapter, but appears in Plass et al. (2011c).

#### 31.4.2.3 Library of Learning Mechanics

To provide learning game designers with a set of learning mechanics and associated instances of game mechanics that are useful for their game designs, we have begun to compile a library of mechanics. This library includes a variety of game mechanics options for each learning mechanic; see Table 31.4 for examples and http://g4li.org for updates.

Note that since learning mechanics are meta-mechanics, there is a one-to-many relationship of learning mechanics to corresponding game mechanics, and each of the different game mechanics that instantiates a Learning Mechanic may only be suitable under specific conditions–for specific learners, topics, or game genres. Our ongoing work is concerned with adding new learning mechanics and associated game mechanics, and with demonstrating their viability and usefulness through empirical research.

### 31.4.3 Assessment Mechanics

In addition to engaging learners in meaningful activities that facilitate learning and assist in the creation of mental models, games have the ability to provide educators and designers, as well as players, with insight into players' learning processes and advancements. The rule systems created by game mechanics can be used to assess a variety of variables, including learning outcomes.

Learning objectives and learning processes of interest can be operationalized into specific actions within the game in ways that allow us to assess achievement levels. Player actions can be captured within the game telemetry, as log files, and can be analyzed to reveal what, and how, players have learnt. For example, different problems in the game might each contribute to a score describing the player's mastery of a particular sub-skill. Game mechanics for assessment must therefore be designed to elicit relevant behaviors that can be observed through the user log and interpreted to reveal learning process, outcome, and learner variables. We call mechanics designed for this purpose *assessment mechanics*.

> Assessment Mechanics are patterns of behavior or building blocks of diagnostic interactivity, which may be a single action or a set of interrelated actions that form the essential diagnostic activity that is repeated throughout a game.

Similar to learning mechanics, assessment mechanics are meta-mechanics that describe player actions but are not playable mechanics. They describe the functions of the actions available to players to demonstrate their knowledge, skills or expressions of other variables of interest, but they don't describe the assessment tasks themselves. For example, the assessment mechanic might specify that the game should engage the players in activities in which they group related items in time or space, but does not describe the corresponding game mechanics, e.g., whether the grouping is done by shifting items on the screen like in *Bejeweled*, dropping them in specific locations like in *Drop Seven*, or placing them like in a *tower defense* game.

In our example of *Angry Birds* used in the Learning Mechanics section, the assessment mechanics may describe are a set of actions that allow us to determine to what extent players have learnt to match properties of the different types of birds (black, green, blue) to properties of the different materials (glass, wood, stone). In this particular case, succeeding in demolishing a structure with a single shot can be interpreted as evidence for a high level of achievement of this learning goal. However, this performance is also indication that players have a good understanding of what trajectory and which force need to be applied to a certain bird, which confounds conclusions about individual outcomes, such as learning to match birds and materials. The assessment mechanic would therefore ask for the implementation of multiple levels with varying problems that allow us to disentangle the different player competencies.

Several variables are of interest in designing personalized or adaptive games (see Table 31.1, above). As discussed above, these variables can be grouped as general trait, general state, and situation-specific state variables. Some of these variables can be reliably assessed with valid traditional instruments, but for many variables

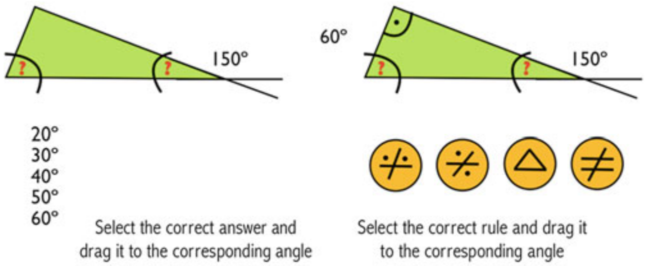


**Fig. 31.2** Two assessment mechanic options in Noobs vs. Leets

only methods with low reliably and validity are available, often involving self-reported data, which are susceptible to learner biases and other response sets. Each variable shown in Table 31.1 is important because it predicts learning and is therefore useful to be measured in-game.

For example, learners' self-regulation is of interest, because it describes whether players establish learning goals, monitor their goal achievement, and change strategies when they are unable to achieve their goals. Our first case example (above) showed how we measured self-regulation in science simulations based on a Cognitive Task Analysis.

Another example is the assessment of specific aspects of learning. In the game *Noobs vs. Leets* (see our third case example below), we were interested in understanding how well the learner comprehends rules related to angles in triangles and quadrilaterals, such as the complementary angles, and opposite angles rules. We, therefore, chose an assessment mechanic requiring learners to drag the correct rule to the angle to be solved (Fig. 31.2, right). An alternative assessment mechanic, requiring the learner to drag or enter the correct numeric value for each angle (Fig. 31.2, left) could not have revealed the source of possible errors, which could have been conceptual (lack of rule knowledge) or arithmetic (lack of subtraction skills). This reasoning is based on the Evidence-Centered Design (ECD) Framework (Mislevy et al. 2003).

This framework, described in more detail above, provides a formal approach to the essential questions related to assessment design: what should be assessed; what kinds of learner behaviors can be used to reveal these constructs, and, what tasks and activities can be designed to elicit these behaviors. Below we will use the ECD model to develop criteria that mechanics must meet to be useful assessment mechanics.

#### 31.4.3.1 Criteria for Assessment Mechanics

Similar to learning mechanics, assessment mechanics must meet criteria to assure they engage the player in meaningful and valid assessment activities. The goal of assessment mechanics is to elicit relevant behaviors that can be observed through the user log and interpreted to reveal learning processes, learning outcomes, and learner variables. To be useful for this purpose, i.e., to make metrics a valid measure producing reliable scores, the first criterion is that assessment mechanics must be based on assessment models, such as ECD. Based on the student model of target competencies, described in relation to gains in skills, knowledge, identity, values, and epistemology that students are supposed to obtain (Rupp et al. 2010), assessment designers need to construct an ECD Evidence Model. Thus, they must specify the salient features of learner behavior and the rules for scoring and interpreting these features for assessment purposes. This involves another important criterion for designing assessment mechanics: the consideration of test-theoretical concerns. For example, since we cannot assume that individual test items are independent of one another, the assessment's statistical model must reflect these possible dependencies (Rupp et al. 2010).

Another criterion is that the mechanics need to be designed based on aspects of the ECD Action Model, i.e., the description of the key tasks and activities in which the learner will engage within the assessment's purpose. These tasks form the assessment mechanic, and it is essential that they be designed so that the task execution can be captured through the game's instrumentation. This means, for example, that the mechanics require the learner to make explicit the steps used for problem solving rather than simply provide the answer to the problem. The mechanics need to create repeated exposures to similar problems to allow for multiple observations of the target behavior.

Depending on the designers' decisions, the character of the assessment mechanic may or may not be obvious to learner. We describe assessments in which learners are aware that they are being assessed as *embedded assessment*, and those where they are unaware of this fact as *stealth assessment* (Shute 2011).

Once measurement experts have designed an assessment mechanic, game designers can design corresponding instances of game mechanics. However, in this process, several design requirements must be met.

#### 31.4.3.2 Requirements for Designing Game Mechanics Based on Assessment Mechanics

One of the most common problems in designing game mechanics based on assessment mechanics is the introduction of confounds that make it difficult to determine whether variability in learning scores among learners can be attributed to their different knowledge and skills or to other factors. One such confound is the addition of new information to be processed and tasks to be executed. For example, game mechanics, such as in *Flight Control*, where players have to determine the approach patterns of airplanes for landing, are fun and engaging because the fast succession of a high number of planes to land puts high demands on players' processing. This could appropriately assess speed of processing, but not conceptual knowledge or higher level thinking. A related confound is the addition of scaffolding or guidance that reduces cognitive task demands for some learners, but not for others. For example, if key information in an adventure game is hidden,

**Table 31.5** Library of assessment mechanics and associated game mechanics

| Assessment mechanic | Corresponding game mechanics |
|---|---|
| Learners apply rules to solve problems | Fling mechanic in *Angry Birds*<br>Drag mechanic in *Explode!*<br>Rule mechanic in *Noobs v. Leets* |
| Learners arrange items in time or space to solve problems | Drag mechanic in *Gravity*<br>Flight path mechanic in *Flight Control*<br>Tile placement mechanic in *Plumber/Toobz* |
| Learners select items that belong to each other in time or space to solve problems | Drag mechanic in *Osmosis*<br>Selection mechanic in *Bejeweled* |

resulting in only some players finding it, then assessment of knowledge will be confounded by learners' exploration strategies and player type.

Another typical problem for assessment is that the game mechanic introduces confounds through demands on fine motor skills, which are highly variable in learners. For example, *MotionMath* asks learners to tilt their tablet device to direct a ball to the correct answer. Success in this task not only depends on learners' knowledge, but also on their ability to move the ball to the correct location.

Likewise, many game mechanics include activities that require learners to have content knowledge or skills from unrelated subject matter areas. For example, an assessment of algebra may be confounded by the need to know about Newtonian physics. Although integrating different subject matter areas is a desirable design feature for learning mechanics, doing so in designing assessment mechanics may confound results.

A final confounding variable to consider is emotion. During game play, learners will likely experience a series of emotional responses that would impact learning outcomes (Um et al. 2011). Designers of assessment mechanics must consider learners' emotions and design mechanics that are aimed at assessment so that the learners' emotional response does not interfere with their ability to solve the problems presented. A particularly problematic situation would result from mechanics in which different people emotionally respond in different ways.

#### 31.4.3.3 Library of Assessment Mechanics

To provide learning game designers with a set of assessment mechanics and associated game mechanics that they can use for their game designs, we have begun to compile a library of mechanics. This library lists a variety of game mechanics options for each assessment mechanic, see Table 31.5 for examples and http://g4li.org for updates.

Note that there is a one-to-many relationship of assessment mechanics to game mechanics, and that each of the different game mechanics that can instantiate an Assessment Mechanic may only be suitable under specific conditions. Our ongoing work is concerned with adding new assessment mechanics and associated game mechanics, and with demonstrating their viability and usefulness through empirical research.

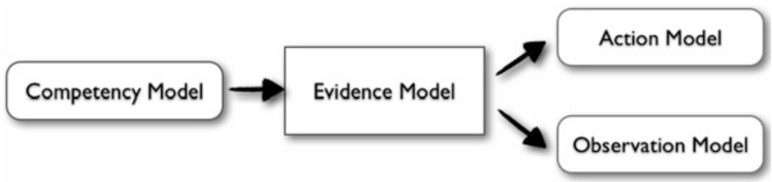


**Fig. 31.3** Expanded ECD framework with added observational model

### *31.4.4 Expanding the ECD Model to Include Observational Measures*

The current ECD model describes how the implementation and analysis of game metrics is based on a systematic process of describing the competencies to be acquired by learners in a competency model. It is crucial to specify the evidence that needs to be collected to be able to assess the degree of competency achieved by a student, and then define an action model that describes learners' actions, characteristics and contextual information. We have extended this framework to allow for additional evidence that goes beyond this strictly behavioral model by incorporating observational measures (see Fig. 31.3).

Observational measures are operationally defined as measures allowing observation of the learner through a variety of measures, including video observations, classroom observers, eye tracking, and biometric sensors such as EKG, GSR, EMG, and EEG.

Adding observational measures to the ECD model expands the evidence model to include new forms of evidence to describe a specific competency or learner variable. Using observational data and data provided by game metrics as evidence of a specific target performance allows triangulation of findings to enhance measurement validity (see Part V "Mixed Methods for Game Evaluation" and Chap. 26). In fact, many of the variables of interest discussed above (see Table 31.1) have correlates in physiological responses. For example, learners' cognitive load can be assessed through behavioral measures of problem solving or other learning tasks, but also through measures of pupil dilation, heart rate, galvanic skin response, and EEG (Isbister and Schaffer 2008). In the following section, we will describe two case examples of how we use this expanded ECD model in our research on games and learning.

## 31.5 Case Example: *FactorReactor*

*FactorReactor* is a G4LI-developed game that focuses on improving math fluency and cognitive flexibility for number sense. Players must complete a series of math problems, but they have a choice of what operations they perform, and on what numbers.

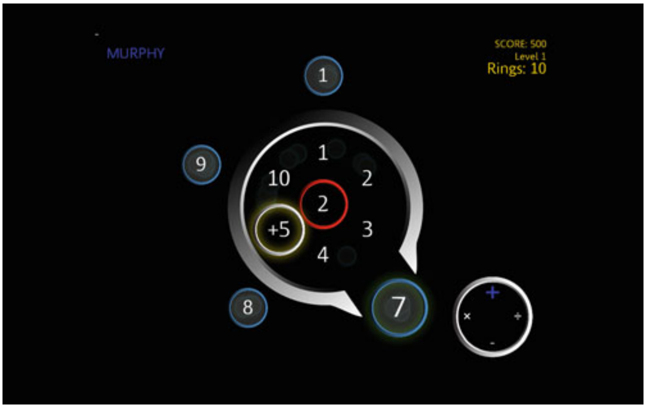


**Fig. 31.4** Screen shot of the G4LI Game *FactorReactor*

### *31.5.1 Game Description*

The Learning Mechanic in *FactorReactor* involves choosing mathematical operations and a second number to change the player's number to the goal number.

The game mechanic instantiating this learning mechanic places the number that the player is trying to manipulate at the center of the screen (Fig. 31.4). The numbers adjacent to the center number and inside the silver circle are the numbers that players can use to change their center number. The numbers outside the silver ring are the players' goals. The object of the game is to make the center number equal each of the goal numbers in the outermost circle.

Players use the Xbox 360's analog stick controller to move the selector circle to the number that they want to use. Mathematical operations (plus, minus, divide, multiply) are mapped to the controller's colored buttons. To perform an operation players pull the right trigger, and to swap the selected number with the center number players pull the left trigger. Also, players can change what goal number they are attempting by pressing the left and right shoulder buttons (Fig. 31.5).

Feedback in the game consists of a point system and a resource system. For the point system, players are rewarded with points for every goal they solve. The resource system is conveyed as rings. Each operation players performs costs 1 ring, and they are rewarded with 5 rings for every goal that is solved. Learners choose what mathematical operations to perform and in what order to implement them. Additionally, since each operation uses one ring, players must be efficient in solving problems occasionally using division and multiplication, or they will run out of rings and have to restart the entire level.

*FactorReactor*'s game mechanics are simple, to reduce players' cognitive load and instantiate the learning mechanic without adding excessive amounts of extraneous load. This allows players to devote more resources to learning activities instead of memorizing what buttons they should press or how to perform an action in the game. Other options, perhaps involving animals or magic spells, could have been


**Fig. 31.5** Mapping of controls for the G4LI Game *FactorReactor*

used, but math fluency is about speed and accuracy, so creating an easy to learn control scheme and simple representations allow players to quickly move from one math operation to the next.

## *31.5.2 Assessment Mechanics*

As discussed above, assessment mechanics describe building blocks of diagnostic activity. In many games, the game mechanics used to instantiate the assessment mechanic are the same as the game mechanics used to instantiate learning mechanics. *FactorReactor* is an example of such a game. The goal of the assessment was to first determine players' initial ability level with mathematical operations, and then to measure how players improved their use of those operations and the speed at which they solved equations.

We conducted a study with *FactorReactor* to determine which mode of play was most effective for increasing math fluency. Sixty-four 6th to 8th grade students were assigned to three versions of the game: *solo*, *collaborative*, and *competitive* (Plass et al. in press). In the solo condition, each student played the game by him or herself. In the collaborative and competitive game versions, students played in pairs, either helping each other or trying to beat each other's score. In preparation, participants first watched a tutorial video and were given a short practice session to become adjusted to the gameplay. In addition to playing the game, participants were asked to complete surveys on situational interest and achievement goal orientation, and pre/post math fluency tests. They also completed play sessions as pre- and post tests. Game play metrics were saved as log files.

**Table 31.6** Selected log data fields captured in *Factor Reactor* game

| |
|---|
| *General information* |
| User ID |
| Game version |
| Treatment |
| *Learning and performance indicators* |
| Total # of levels completed |
| Total # of problems solved |
| Total # of unique problems solved |
| Total # of rings |
| Total score |
| Total # of "+" operations used per solution |
| Total # of "–" operations used per solution |
| Total # of "x" operations used per solution |
| Total # of "÷"operations used per solution |
| Total # of operators used per solution |
| *Self-Regulation Patterns* |
| Total # of game overs |
| Total # of operations used per solution |
| Total # of operations used per game over |
| *Cognitive Flexibility* |
| Total # of "swap" operations used |
| Total # of "x" operations used per solution |
| Total # of "÷"operations used per solution |

### *31.5.3 Game Telemetry*

*FactorReactor* was instrumented to capture in-game player behavior in a log file. The assessment mechanic was designed to allow us to capture the specific variables within the game that would provide evidence of what rules players were selecting, how players implemented the rules and in what order, and if the players were successful. Perhaps the simplest question to answer with the telemetry data was what operators players were using. We stored each player's action, which allowed us to count all additions, subtractions, multiplications, and divisions players performed on each attempt at a level. Thus, if players failed and restarted a level, we could distinguish between what they did on their first attempt and how their strategy may have changed with each subsequent attempt (Table 31.6).

The frequency of each operation provides insight into players' ability level and their comfort with each operation. For instance, our study of middle school students revealed that many participants made infrequent use of multiplication and division, preferring to stick with simpler operations of addition and subtraction.

Other telemetry data that were useful in understanding player performance included time taken to solve a goal and to solve a level, number of attempts per level, and rings used per attempt. For example, an analysis of the overall number of problems solved and controlling covariates, such as pre-test game performance and degree of experience with a video game controller, found that players in the competitive condition performed significantly better than players in the individual condition, but not significantly better than those in the collaborative condition.

The metric of time taken can offer a direct assessment of ability level in situations where the demands of processing unrelated information and executing unrelated tasks is minimized. For instance, in cases where control schemes are difficult to master or confusing, time taken may only reflect the players' struggles to input their knowledge. This can be assessed through the use of pre- and post-tests of the educational content delivered in a more traditional way, such as a paper and pencil assessment. Time-stamped user actions can also be analyzed in other ways. For instance, we can mark the time each operation was used in the log files to determine what order players used for operations when solving a goal or level. This can indicate players' strategies, such as using the common concept of order of operations.

We did not assess players' cognitive flexibility with a paper-based measure, but the user logs allowed us to analyze what operations were used, how many swaps were performed, how often the player changed goals, and the time taken to solve a goal. Each of these variables can be analyzed based on each attempt to solve a problem, level, or overall. Players with higher cognitive flexibility for the task are likely to solve problems faster, use more diverse operations, and to use swapping and goal switching.

### *31.5.4 Future Directions*

In addition to the metrics discussed, psychophysiological measures can provide rich sources of information on players. With *FactorReactor*, we have begun collecting data from participants using an eyetracker. Eyetracking data allows us to mark areas of interest on the gameplay screen and analyze a variety of different variables related to the players' gaze. For example, we can measure the frequency that players look at a specific area of the screen, or how long they spend fixated on that area. We can also analyze areas of interest that players commonly transition between. In our *Factor Reactor* study, for example, we can compare the competitive condition with the other conditions and determine if players spent more time or glanced more frequently at their score or the score of their opponents. Eyetracking data can provide information about player strategy, player motivation, and resource management. Eyetracking can also provide data regarding extraneous cognitive load, if it shows that players frequently glance at areas of the screen containing information that is not useful for game play.

## 31.6 Case Example: *Noobs vs. Leets*

Our final case example describes *Noobs vs. Leets*, another game developed at G4LI, designed to teach middle school geometry. More specifically, players learn about different types of angles and rules that can be used to determine missing angles in quadrilaterals.

### *31.6.1 Game Description*

*Noobs vs. Leets* uses a simple narrative in which players must guide their characters, noobs, across levels to save their fellow noobs from their dreaded enemy, the leets. Through identifying angles, players open up pathways that allow them to reach their goal. The game is divided into six chapters, each with about eight levels. In each chapter students learn about a different angle type.

Chapter 1 starts with simple angles. As players progress, they encounter complementary angles, supplementary angles, vertical angles, triangles, and finally quadrilaterals. The beginning of each chapter includes a cut-scene that introduces the new angle type.

Figure 31.6 provides an example of a typical level midway through the game. Each level starts with the noob parachuting in from an airplane and starting at the

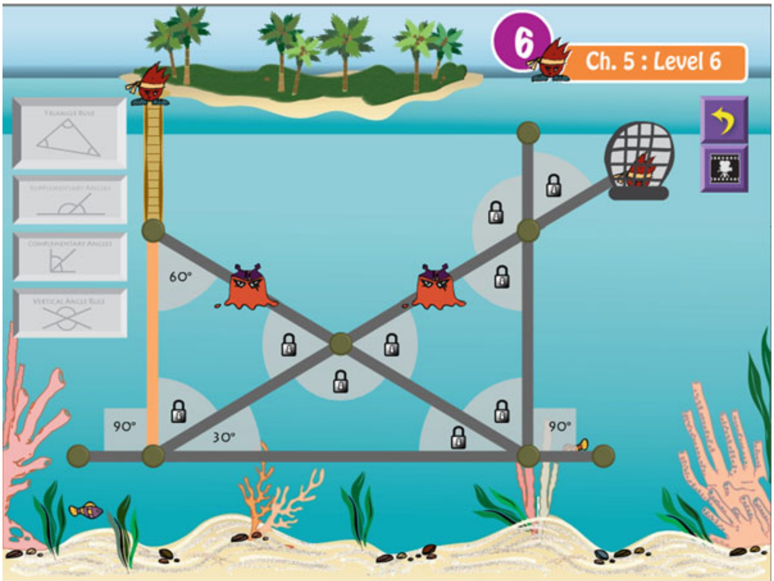


**Fig. 31.6** Chapter 5 of the G4LI Game *Noobs v. Leets*

upper left portion of the screen. By using known angles, players can solve locked angles by applying the correct rule. On the left side of the screen are four buttons, corresponding to different types of angles. As players progress through the game, they learn about and apply new and more complex rules. Each time a player unlocks an angle, a new pathway opens. The goal is for players to reach the noob trapped in the cage in order to set it free. On some levels, there are stationary or mobile leets that try to knock the player off the level and force them to restart the level from the beginning. Each time a level is solved, the player earns another noob; however, if they choose the wrong angle or run into a leet, they lose a noob. The number of noobs a player has saved is indicated next to the chapter/level indicator. On the right side of the screen, players have access to a button that allows them to move to a previous level or re-watch the cut-scene for that chapter.

### *31.6.2 Learning Mechanics*

*Noobs vs. Leets* provides a useful example of how learning mechanics can shape a game mechanic. The learning mechanic is *Apply Rules to solve Problems: Learner selects among different rules and indicates for which problems they apply.*

Although *FactorReactor* and *Noobs vs. Leets* look quite different and have dissimilar play, there are similarities in their underlying learning mechanics. In both games, players have to choose the rule (angles or mathematical operations) and apply it where they see fit. One difference is that rules are of a more conceptual level in *Noobs vs. Leets*. To assist learning, the first few levels of each chapter provide scaffolding to establish a foundational knowledge for that new rule. As the player progresses through each chapter, levels shift from practice and scaffolds to problem solving. Players must decide not only what rule is most applicable, but also what route is the best to take through the level.

### *31.6.3 Assessment Mechanics*

The assessment mechanics for *Noobs vs. Leets* was designed to allow us to decide if players understood the rules they were applying, how successful they were at implementing the rules, and if they were able to improve as they progressed in the game. For example, we were interested in gathering evidence about players' ability to grasp the concept of complementary angles and whether they were able to apply it with fewer mistakes as they finished each level.

We conducted a study with 89, 6th to 8th grade participants to assess if it was more important for players to apply the rule to an angle (as described above) or if they would learn more from calculating the angle (Plass et al. 2012). For this purpose we created two versions of *Noobs vs. Leets*. One worked as described above; the second version included number buttons instead of rule buttons and players had

**Table 31.7** Selected log data fields captured in *Noobs v. Leets* game

| |
|---|
| **General information** |
| User ID |
| Game version |
| Treatment |
| **Learning and performance indicators** |
| Total # of levels completed |
| Total # of unique correct solutions |
| Total # of unique correct solutions per chapter |
| Total # of incorrect solutions |
| Total # of incorrect solutions per chapter |
| Total play time |
| Time per chapter |
| Total # of noobs lost |
| Total # of noobs lost per chapter |
| Total "Game over" screens |
| Total clicks on vertices |
| Total clicks on angles |
| Total # of unique correct uses for each rule |
| Total # of incorrect uses for each rule |
| Time on each tutorial cut-scenes |
| Total time on tutorial cut-scenes |
| Total number of times player returned to the tutorial cut-scene for each chapter |
| Average reaction time to level per chapter |
| Total # of deaths to leets |
| **Self-regulation patterns** |
| Average # of consecutive incorrect rule applications in a row using the same rule |
| Longest string of consecutive incorrect uses of each rule |
| **Cognitive flexibility** |
| Average number of different rules used per level |
| Highest number of different rules used on a level |

to calculate the locked angles and then click on the correct number of degrees for that angle rather than selecting the rule that they would apply. The difference between the two treatments, therefore, was that one explicitly captured which rule players applied, whereas the other only captured the numeric responses for solved angles. Participants played the game for 25 min and completed a situational interest survey, a math fluency test, and a pretest/posttest of geometry knowledge. The math fluency test and pretest allowed us to run analyses controlling for prior knowledge.

### *31.6.4 Game Telemetry*

For this study, we were interested in capturing telemetry data in the log files that would allow us to compare our two conditions in several ways. Table 31.7

summarizes variables that were tracked in the log files. These variables were assessed to measure general information, learning and performance indicators, self-regulation, and cognitive flexibility.

Telemetry data paired with surveys and pre- and posttest measures allowed us to analyze the data from several perspectives. For example, results indicated that although players in the number condition had greater situational interest in the game, players in the rules condition completed many more levels of the game. As for learning outcomes, we found that players who used the rule-based game continued increasing their knowledge as they completed more levels, whereas players of the number-based version did not increase their knowledge past a certain level. Results of players in the conceptual or rule condition did not exhibit an interaction between test performance and levels completed. For a more detailed analysis and additional results, see Plass et al. (2012).

### *31.6.5 Future Directions*

In our next iteration of research using *Noobs v. Leets*, we are incorporating an incentive system that uses badges as rewards for different accomplishments in the game. The badges are designed based on goal orientation theory, with some badges appealing to a mastery orientation, while others appeal to a performance orientation (Ames and Archer 1988; Elliot 2005). For this study, we will be instrumenting the log files to show how many badges players earned and how often they clicked to view their badges. In addition, we will use eyetracking data to determine if there is some optimal level of badge implementation. Are there cases where players look at badges too much and become distracted from the learning objectives? Or do the badges provide sufficient motivation to improve learning?

Additionally, work is underway examining how the two versions of the game, the conceptual and the numeric, interact with student engagement and prior knowledge. Preliminary results suggest students with higher ability beliefs in math enjoy solving problems numerically, whereas students with lower ability beliefs tend to prefer the rule based system. Such work has implications for how learning mechanics are integrated into games for learning. Work in this area is also examining motivation over time within the game, and the effect of varying choice and feedback within assessment mechanics.

## 31.7 Summary and Conclusion

This chapter introduced the approach for assessment of learning and related learner variables taken by the Games for Learning Institute (G4LI), with a particular focus on the use of metrics obtained during game play. Our approach suggests that game mechanics, the essential game play activity, should be distinguished from learning mechanics and assessment mechanics. We define learning mechanics as essential high-level activities, grounded in learning sciences; they have learning as the primary objective. In contrast, assessment mechanics are essential high-level activities, grounded in test theory, that have assessment as the primary objective. Learning and assessment mechanics are meta-mechanics that can be instantiated as corresponding game mechanics, following criteria we outlined above to preserve their intended teaching or assessment objective.

Variables related to learning that can be measured through game metrics include learning outcomes (cognitive and skills), trait variables, general state variables, and situation-specific state variables. Supplementing log data of game events and user behavior with observational data extends the ECD model and can result in more valid assessments of these variables.

Our approach serves two related but separate goals. One is a measurement goal–embedded assessment allows for more detailed insights into learning than many traditional instruments, both with respect to the process of learning and learning outcomes. This has implications for research as well as learner competency testing. The other goal is related to improving game play. By using assessment mechanics to measure a series of learner variables, a competency model can be compiled, allowing for the design of games that are individualized and adaptive to a learner's specific needs and characteristics. This has implications for the design of effective games for learning by making games more adaptive and personalized, and, hopefully, more effective.

Using this approach, designers of games for learning have a strategy at their disposal that allows them to design the mechanics of their game based on the specific goal of either learning or assessment of learning, or both. The resulting learning and assessment mechanics communicate these functions to game designers and provide corresponding constraints for the design of game mechanics. For each game and game genre these mechanics will differ, and libraries of mechanics, such as the ones under development by G4LI, will provide useful resources for learning game designers.

In summary, we described an approach that, grounded in theory and tested in several game design projects, has implications both for research and practice of the design of games for learning. While many other approaches have applied learning theory and assessment theory to game design, the approach presented here provides a practical way for teams designing games for learning to separate the role of learning scientists (designers of learning mechanics), assessment experts (designers of assessment mechanics), and game designers (designers of the corresponding game mechanics).

## About the Authors

**Jan L. Plass, Ph.D.** is Paulette Goddard Professor of Digital Media and Learning Sciences at NYU Steinhardt. He is the founding director of the Consortium for Research and Evaluation of Advanced Technology in Education (CREATE), and co-director of the Games for Learning Institute (G4LI).

**Bruce D. Homer, Ph.D.** is an Associate Professor of Educational Psychology (Learning, Development & Instruction Sub-Program) and Training Director for the

Interdisciplinary Postdoctoral Research Training program at the Graduate Center at the City University of New York (CUNY).

**Charles K. Kinzer, Ph.D.** is a Professor of Communication and Education at Teachers College Columbia University where he is the program coordinator at Computing, Communication and Technology in Education (CCTE).

**Yoo Kyung Chang, Ph.D.** is a Lecturer in CCTE at Teachers College Columbia University and former research assistant at the CREATE lab at New York University.

**Jonathan Frye** is the Technology Coordinator and a Research Assistant for the CREATE lab. He is currently a doctoral candidate in the Educational Communications and Technology program at New York University.

**Walter Kaczetow** is currently a doctoral student in CUNY Graduate Center's program in Educational Psychology. When not studying Walter can be found teaching mathematics as an adjunct professor at New Jersey City University.

**Katherine Isbister, Ph.D.** is an Associate Professor of both Computer Science Engineering and Digital Media at NYU's Polytechnic Institute. She is the founding director of the Social Game Lab.

**Ken Perlin, Ph.D.** is a professor in the Department of Computer Science at New York University. In addition to being the director of the Games For Learning Institute, he was also founding director of the Media Research Laboratory and director of the NYU Center for Advanced Technology.

## Recommended Readings


Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010). *Automatic assessment of complex task performance in games and simulations* (CRESST Report 775). Los Angeles: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kim, J. H., Gunn, D. V., Schuh, E., Phillips, B., Pagulayan, R. J., & Wixon, D. (2008). Tracking real-time user experience (TRUE): A comprehensive instrumentation solution for complex systems. In *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 443–452). Florence: ACM. doi:10.1145/1357054.1357126

Plass, J. L., O'Keefe, P., Homer, B. D., Case, J., Hayward, E. O., Stein, M., & Perlin, K. (in press). Motivational and cognitive outcomes associated with individual, competitive, and collaborative game play. Special issue on advanced learning technologies. *Journal of Educational Psychology*.

Plass, J. L., Homer, B. D., Kinzer, C., Frye, J., & Perlin, K. (2011, September 30). *Learning mechanics and assessment mechanics for games for learning* (G4LI White Paper # 01/2011 Version 0.1). Available online at www.g4li.org

Shute, V. J., Ventura, M., Bauer, M. I., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow. In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (pp. 295–321). Mahwah: Routledge/Taylor & Francis.

## References


Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In Gauthier, G., Frasson, C., & VanLehn, K. (Eds.), *Proceedings of the 5th international conference on intelligent tutoring systems, ITS 2000* (pp. 292–303). Berlin: Springer.

Ambinder, M. (2009). *Valve's approach to playtesting: The application of empiricism*. Presented at the 2009 Game Developers Conference, San Francisco.

America's Army. (2012). *America's Army: Honor and advances – Army values*. Retrieved July 3, 2012 from http://manual.americasarmy.com/index.php/Honor_and_Advancement

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80*(3), 260–267.

Anderson, A., & Bavelier, D. (2011). Action game play as a tool to enhance perception, attention and cognition. In S. Tobias & J. D. Fleycher (Eds.), *Computer games and instruction* (pp. 307–330). Charlotte: Information Age Publishing: IAP.

Brown, T. L. (1997). *Task analysis strategies and practices* (Practices Application Brief). Washington, DC: Office of Educational Research and Improvement. (ERIC Document Reproduction Service No. ED 404 571).

Chang, Y. K. (2010). *Examining metacognitive processes in exploratory computer-based learning environments using activity log analysis*. Unpublished Doctoral Dissertation, New York University, New York.

Chang, Y. K., & Plass, J. L. (2012). Assessment of the metacognitive processes from the behavioral data. (G4LI White Paper # 01/2012). Available online at www.g4li.org

Chang, Y. K., Plass, J. L., & Homer, B. D. (2008, October). *Development and validation of a behavioral measure of metacognitive processes (BMMP)*. Featured research presentation at the annual convention of the Association for Educational Communication and Technology (AECT), Orlando, FL.

Cognition and Technology Group at Vanderbilt. (1990). Anchored instruction and its relationship to situated cognition. *Educational Researcher, 19*, 2–10.

Cognition and Technology Group at Vanderbilt. (1993). Anchored instruction and situated cognition revisited. *Educational Technology, 33*(3), 52–70.

Collins, A. (1988). *Cognitive apprenticeship and instructional technology*. Hillsdale: Lawrence Erlbaum.

Cook, D. (2006). What are game mechanics? *lostgarden.com.* Retrieved May 23, 2010 from http://lostgarden.com/2006/10/what-are-game-mechanics.html

Cordingley, E. S. (1989). Knowledge elicitation techniques for knowledge-based systems. In D. Diaper (Ed.), *Knowledge elicitation: Principles, techniques, and application* (pp. 89–175). New York: Wiley.

DeRosa, P. (2007, August 7). Tracking player feedback to improve gamedesign. *Gamasutra*.

Domagk, S., Schwartz, R., & Plass, J. L. (2010). Defining interactivity in multimedia learning. *Computers in Human Behavior, 26*, 1024–1033. doi:10.1016/j.chb.2010.03.003.

Ducheneaut, N., Moore, R. J., & Nickell, E. (2004). Designing for sociability in massively multiplayer games: An examination of the "third places" of SWG. In J. H. Smith & M. Sicart (Eds.), Proceedings of the Other Players Conference. Copenhagen, Denmark.

Efklides, A. (2002). They systemic nature of metacognitive experiences: Feelings, judgment, and their interrelation. In M. Izaute, P. Chambres, & P.-J. Marescaux (Eds.), *Metacognition: Process, function, and use* (pp. 19–34). Dordrecht: Kluwer.

Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52–72). New York: Guilford Publications.

Gee, J. P. (2008). *What video games have to teach us about learning and literacy* (Rev. and updated). Basingstoke: Palgrave Macmillan.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–38.

Heeter, C., Magerko, B., Medler, B., & Fitzgerald, J. (2009). Game design and the challenge-avoiding 'Validator' player type. *International Journal of Gaming and Computer-Mediated Simulations, 1*(3), 53–67.

Hill, J. R., & Hannafin, M. J. (2001). Teaching and learning in digital environments: The resurgence of resource-based learning. *Educational Technology Research and Development, 49*, 15–26.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the Challenges in Game AI Workshop, 19th National Conference on Artificial Intelligence,* AAAI'04, San Jose, CA. Vancouver: AAAI Press.

Isbister, K., & Schaffer, N. (2008). *Game usability. Advice from the experts for advancing the payer experience*. New York: Morgan Kaufman.

Isbister, K., Flanagan, M., & Hash, C. (2010). Designing games for learning: Insights from conversations with designers. In *Proceedings of CHI* (*Conference on human factors in computing*), Atlanta, GA.

Järvinen, A. (2008). *Games without frontiers: Theories and methods for game studies and design*. Tampere: Tampere University Press.

Juul, J. (2003). The game, the player, the world: Looking for a heart of gameness. In M. Copier & J. Raessens (Eds.), *Level up: Digital games research conference proceedings* (pp. 30–45). Utrecht: Utrecht University.

Lajoie, S. P., Azevedo, R., & Fleiszer, D. M. (1998). Cognitive tools for assessment and learning in a high flow information environment. *Journal of Educational Computing Research, 18*(3), 205–235.

Lave, J., & Wenger, E. (1990). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Liu, M. (1998). A study of engaging high-school students as multimedia designers in a cognitive apprenticeship-style learning environment. *Computers in Human Behavior, 14*(3), 387–415.

McCombs, B. L. (2001). Self-regulated learning and academic achievement: A phenomenological view. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 67–123). Hillsdale: Lawrence Erlbaum Associates.

Mislevy, R. J., & Gitomer, D. H. (1996). The role of probability-based inference in an intelligent tutoring system. *User-Modeling and User-Adapted Interaction, 5*, 253–282.

Mislevy, R. J., & Riconscente, M. (2005). *Evidence-centered assessment design: Layers, structures, and terminology* (PADI Technical Report 9). Menlo Park: SRI International.

Mislevy, R. J., Steinberg, L. S., Breyer, F. J., Almond, R. G., & Johnson, L. (1999). A cognitive task analysis with implications for designing a simulation-based assessment system. *Computers in Human Behavior, 15*, 335–374.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language Testing, 19*, 477–496.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives, 1*, 3–67.

Nelson, B., Erlandson, B., & Denham, A. (2010). Global channels for learning and assessment in complex game environments. *British Journal of Educational Technology.* Published online, January 2010. To appear in print, June 2010.

Perry, N. E., VandeKamp, K. O., Mercer, L. K., & Nordby, C. J. (2002). Investigating teacher-student interactions that foster self-regulated learning. *Educational Psychologist, 37*(1), 5–15.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ). *Educational and Psychological Measurement, 53*, 801–813.

Plass, J. L., O'Keefe, P., Homer, B. D., Case, J., Hayward, E. O., Stein. M., & Perlin, K. (in press). Motivational and educational outcomes associated with individual, competitive, and collaborative game play. *Journal of Educational Psychology*.

Plass, J. L., Homer, B. D., Hayward, E. O., Frye, J., Huang, T. T., Biles, M., Stein, M., & Perlin, K. (2012, September 18–20). The effect of learning mechanics design on learning outcomes in a computer-based geometry game. *Paper presented at GameDays 2012* held, at Fraunhofer IGD TU Darmstadt, Darmstadt, Germany.

Roth, E. M., & Woods, D. D. (1989). Cognitive task analysis: An approach to knowledge acquisition of intelligent system design. In G. Guida & C. Tasso (Eds.), *Topics in expert system design* (pp. 233–264). New York: Elsevier Science.

Rupp, A. A., Gushta, M., Mislevy, R. J., & Shaffer, D. W. (2010). Evidence-centered design of epistemic games: Measurement principles for complex learning environments. *Journal of Technology, Learning, and Assessment, 8*(4). Retrieved June 1, 2011 from /http://www.jtla.org

Salen, K., & Zimmerman, E. (2003). *Rules of play: Game design fundamentals*. Cambridge: MIT Press.

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (Eds.). (2000). *Cognitive task analysis*. Mahwah: Lawrence Erlbaum Associates.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351–371.

Schwartz, D. L., & Black, J. B. (1990). *The induction of rules from analog, mental models*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Shaffer, D. W. (2006). *How computer games help children learn*. New York: Palgrave Macmillan.

Shute, V. J. (2010). *Innovative assessment for the 21st century: Supporting educational needs*. New York: Springer.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 503–524). Charlotte: Information Age.

Shute, V. J., & Kim, Y. J. (2011). Does playing the World of Goo facilitate learning? In D. Y. Dai (Ed.), *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning* (pp. 243–267). New York: Routledge Books.

Shute, V. J., & Torreano, L. A. (2003). Formative evaluation of an automated knowledge elicitation and organization tool. In Murray, T., Ainsworth, S., & Blessing, S. (Eds.), *Authoring tools for advanced technology learning environments: Toward cost-effective adaptive, interactive, and intelligent educational software* (pp. 149-180). The Netherlands: Kluwer Academic Publishers.

Sicart, M. (2008). Defining game mechanics. *Game Studies, 8*(2). Retrieved June 1, 2011 from http://gamestudies.org/0802/articles/sicart

Spiro, R. J., Coulson, R. L., Feltovich, P. J., & Anderson, D. (1988). Cognitive flexibility theory: Advanced knowledge acquisition in ill-structured domains. In V. Patel (Ed.), *Proceedings of the 10th annual conference of the cognitive science society*. Hillsdale: Erlbaum.

Spiro, R. J., & Jehng, J. C. (1990). Cognitive flexibility and hypertext: Theory and technology for the nonlinear and multidimensional traversal of complex subject matter. In Nix, D. & Spiro, R. (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 163–205). Hillsdale: Erlbaum.

Swink, S. (2008). *Game feel: A game designer's guide to virtual sensation*. New York: Morgan Kaufmann.

Tychsen, A., & Canossa, A. (2008). Defining personas in games using metrics. In *Proceedings of the 2008 conference on future play: Research, play, share* (pp. 73–80). New York/Boston: Association Computing Machinery.

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2011). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498.

Williamson, D., Bauer, M., Steinberg, L. S., Mislevy, R. J., Behrens, J., & DeMark, S. (2004). Design rationale for a complex performance assessment. *International Journal of Testing, 4*, 303–332.

Winne, P. H. (1982). Minimizing the black box problem to enhance the validity of theories about instructional effects. *Instructional Science, 11*, 13–28.