

MIT Open Access Articles

*Approaches to illuminate content-specific
gameplay decisions using open-ended game data*

The MIT Faculty has made this article openly available. **Please share**
how this access benefits you. Your story matters.

As Published: <https://doi.org/10.1007/s11423-021-09989-0>

Publisher: Springer US

Persistent URL: <https://hdl.handle.net/1721.1/136830>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of use: Creative Commons Attribution-Noncommercial-Share Alike



Approaches to illuminate content-specific gameplay decisions using open-ended game data

Cite this article as: Louisa Rosenheck, Meng-Tzu Cheng, Chen-Yen Lin and Eric Klopfer, Approaches to illuminate content-specific gameplay decisions using open-ended game data, Educational Technology Research and Development <https://doi.org/10.1007/s11423-021-09989-0>

This Author Accepted Manuscript is a PDF file of an unedited peer-reviewed manuscript that has been accepted for publication but has not been copyedited or corrected. The official version of record that is published in the journal is kept up to date and so may therefore differ from this version.

Terms of use and reuse: academic research for non-commercial purposes, see here for full terms. <https://www.springer.com/aam-terms-v1>

Author accepted manuscript

Approaches to illuminate content-specific gameplay decisions using open-ended game data

Abstract

Games can be rich environments for learning and can elicit evidence of students' conceptual understanding and inquiry processes. Illuminating students' content-specific gameplay decisions, or methods of completing game tasks related to a certain domain, requires a context that is open-ended enough for students to make choices that demonstrate their thinking. Doing this also requires rich log data and methods of Game Learning Analytics (GLA) that are granular enough to look at the specific choices most relevant to that context and domain. This paper presents research done on student exploration of high school level Mendelian genetics in a multiplayer online game called *The Radix Endeavor*. The study uses three approaches to identify content-specific gameplay decisions and distinguish players utilizing different methods, looking at actions and tool use, play patterns and player types, and tool input patterns. In the context of the selected game quest, the three approaches were found to yield insights into different ways that students complete tasks in genetics, suggesting the potential for a set of more generalized guiding questions in the GLA field that could be adopted by learning games designers and data scientists to convey information about content-specific gameplay decisions in learning games.

Keywords: Games; science learning; gameplay patterns; game learning analytics

Author names and affiliations

1. Louisa Rosenheck
Massachusetts Institute of Technology
MIT Education Arcade
Cambridge, Massachusetts 02139
2. Meng-Tzu Cheng (corresponding author)

National Changhua University of Education
Department of Biology
No. 1, Jin-De Road, Changhua 500, Taiwan
Office: +886-4-723-2105 ext.3419
Fax: +886-4-721-1156
Email: mtcheng@cc.ncue.edu.tw

3. Chen-Yen Lin
Eli Lilly and Company
Indianapolis, Indiana 46285 USA
4. Eric Klopfer
Massachusetts Institute of Technology
MIT Education Arcade
Cambridge, Massachusetts 02139

Biography

1. Louisa Rosenheck is a research manager in The Education Arcade at MIT. She manages the design, content, and development of a wide array of educational technology projects and oversees the research done on these projects, exploring how games can be used most effectively in both formal and informal educational settings.
2. Meng-Tzu Cheng is a professor in Department of Biology at National Changhua University of Education in Taiwan. Her research interest mainly focuses on how educational games can be effective learning tools and investigates the impact of the interactions between players and game mechanisms on student science learning through serious gaming.
3. Dr. Chen-Yen Lin is a Sr. Research Scientist with Eli Lilly and Company. He is specialized in applying advanced analytical approaches to solve non-routine medical and health outcome questions.
4. Eric Klopfer is Professor and Director of the Scheller Teacher Education Program and The Education Arcade at MIT. His work uses a design-based research methodology to span the educational technology ecosystem, from design and development of new technologies to professional development and implementation.

Funding

This research was funded by the Bill and Melinda Gates Foundation and the Ministry of Science and Technology (MOST), Taiwan, under grant contract no. 104-2918-I-018-008 and 105-2511-S-018-015-MY5.

Conflict of Interest

The authors declare that they have no conflict of interest.

Author accepted manuscript

Introduction

Serious games, games designed for purposes other than pure entertainment, have many valuable affordances for learning (Ke, 2016). In addition to presenting experiences for learners to build an understanding of content-related concepts and broader skills, they also provide a mechanism by which researchers and educators can gain insight into students' gameplay decisions and approaches. This refers to decisions students make about which tools to use, what game objects to use them on, and what they do with the resulting information or artifacts. It is done by collecting log file data and utilizing Game Learning Analytics (GLA) methods to make sense of the data (Freire et al., 2016; Loh, Sheng, & Ifenthaler, 2015b). In order to provide a context in which students can demonstrate various gameplay decisions, it is necessary to have a context that is open-ended enough to provide authentic choices to players, and analysis methods that look at the content-specific choices they make, rather than more generic achievements in the game. Observing content-specific gameplay decisions, or methods of completing tasks within a certain domain, can lead to insights for teachers to adjust instruction, but first game designers and data scientists must design appropriate metrics relevant to the game's purpose. This paper contributes a set of approaches to GLA that demonstrate how content-specific gameplay decisions can be illuminated in the context of *The Radix Endeavor (Radix)*, a game that was intentionally designed for open-ended exploration of Mendelian genetics. It will describe the process of applying each approach to the specific mechanics of *Radix* to understand how certain types of analyses tease out differences in student gameplay decisions. Based on the results, the paper will then discuss potential positive impacts of applying these approaches to content-specific analysis of other games.

Background

Game Learning Analytics

In the past decade, a rich body of research has illuminated the many ways that digital games can support teaching and learning (Bado, 2019; Qian & Clark, 2016). Given that digital games have the potential for much more than just entertainment, they are now being designed and implemented specifically for education, and these digital games with a purpose beyond entertainment fall into the category of serious games (Loh, Sheng, & Ifenthaler, 2015a). In a serious gaming environment the players can interact with the system in multiple ways, and learning is considered dynamically contextualized within gameplay. As a result, GLA, using data science techniques to analyze multiple types of gameplay data collected from serious games, is growing rapidly, with researchers like Freire et al. (2016) and others creating an infrastructure for the field.

GLA, also known as serious games analytics, is the application of statistical metrics to analyse gameplay data derived from well-designed log files, and collected from a sequence of actions and choices within complex tasks, to provide actionable insights that improve game design and learning performance (Loh et al., 2015; Rowe, Asbell-Clarke, & Baker, 2015). This kind of rich information can let teachers and researchers observe students' approaches to a problem, and then tailor their guidance for each student to improve learning and teaching, as well as contribute to data-driven game design (Alonso-Fernández, Calvo-Morata, Freire, Martínez-Ortiz, & Fernández-Manjón, 2019). Although learning analytics is well established, there remains a lack of understanding of student interactions in games and how the learning process occurs (Freire et al., 2016). Few empirical studies have been conducted that speak to the learning process (Alonso-Fernández et al., 2019), and there is still a need for more research on learning analytics, assessment, and games (Liu, Kang, Liu, Zou, & Hodson, 2017). The lack of

standardization or systematization across GLA applications may be preventing more widespread use across learning games (Alonso-Fernández, Calvo, Freire, Martinez-Ortiz, & Fernandez-Manjon, 2017; Alonso-Fernández et al., 2019) . In addition, while some GLA applications utilize action specific data and data on student choice sequences, many depend on more generic metrics such as task completion, correct question responses, and time spent (Chaudy, Connolly, & Hainey, 2014). One important role of digital data in learning games is to provide feedback on not only what players have learned and can do, but also how they interact with the tool as they engage in the process of learning. Content-specific gameplay decisions are key parts of that process and must be considered in the analysis of learning games data. Given the current state of GLA and its potential for supporting learning, this study aims to contribute to the field by implementing GLA methods on empirical data collected in situ, and proposing a set of guiding questions that can be applied to develop metrics that differentiate students' content-specific gameplay decisions.

Challenges in genetics learning

To set the stage for the content-specific gameplay decisions identified in this study, we start with some context on known challenges for students studying genetics. Genetics is commonly taught yet often challenging for both teachers and students (Bahar, Johnstone, & Hansell, 1999; Duncan & Reiser, 2007; Etobro & Banjoko, 2017; Karagoz & Cakir, 2011; Lewis & Kattmann, 2004). Reasons for this include not only that the structure of the content knowledge is complex, but also that this knowledge should be used to solve complex problems. The complex nature of genetics knowledge lies in the fact that its concepts inherently involve multiple thought levels: the macro (phenotypes of an organism), the micro (cells and chromosomes), the sub-micro/biochemical (DNAs), and the symbolic (genotypes), each with an inseparable connection

to the others (Chu & Reid, 2012; Johnstone, 1991). By definition, a phenotype is an individual's observable trait, and its genetic contribution is the genotype (NHGRI, n.d.). To better construct causal explanations of genetic phenomena from a systems perspective, the different levels of thought must be appropriately integrated. However, simultaneously dealing with and shifting between the levels is no easy task for students (Wilensky & Resnick, 1999), and this confusion of levels generally results in numerous misunderstandings about genetics. Knippels, Waarlo, & Boersma (2005) suggest four design criteria for teaching and learning genetics, one of which is to provide learning activities that allow students to actively explore the relations between those levels.

Problem-solving is essential to genetics. Learning genetics must entail complex problem-solving across the different thought levels (Knippels & Waarlo, 2018), and understanding of genetics is usually assessed by the ability to solve problems. Generally, there are two types of genetics problems, cause-to-effect – reasoning from the known data (genotypes, inheritance patterns) to a specific solution (the distribution of the genotypes and phenotypes among offspring), and effect-to-cause – inferring about inheritance patterns from the observed/desired traits of parents and offspring (Collins & Stewart, 1989; Tsui & Treagust, 2010). The cause-to-effect problems are close-ended ones that can be carried out by applying a set of algorithms mechanically, whereas the effect-to-cause problems are open-ended ones that usually have more than one solution and oblige students to propose hypotheses and test various strategies (Aznar & Orcajo, 2005). Research suggests that open-ended problems work better than close-ended ones, because open-ended problems require students to put their knowledge into play and redefine their conceptual schema when necessary (Orcajo & Aznar, 2005). However, close-ended

problems are common in textbooks and can more easily be taught with conventional methods, resulting in open-ended problems appearing less often in educational settings.

Massively multiplayer online games (MMOGs), a genre of digital games that have been shown to provide authentic learning experiences in which players frequently engage in scientific thinking and inquiry (Ketelhut, 2007; Steinkuehler & Duncan, 2008), have great potential to assess students' understanding of genetics by embedding effect-to-cause problems into a play context in a motivating and non-intrusive way. Typically in a MMOG, a series of quests which set specific goals, are presented to players. The players are often equipped with different tools that can be used to experiment and interact with the virtual world and its inhabitants. They have to propose hypotheses, test strategies, reflect on the results, and repeat these steps to find out how to complete the quests to accomplish goals. Through this inquiry-based experience involving active exploration and discovery, players can construct an understanding of the world's underlying systems (Rosenheck, Gordon-Messer, Clarke-Midura, & Klopfer, 2016).

The Radix Endeavor was therefore developed to capitalize on the affordances of MMOGs (Klopfer, Haas, Osterweil, & Rosenheck, 2018). It was designed to be a more open-ended inquiry environment providing the opportunity to explore in different ways, in which we can capture which traits players are interested in, which organisms they choose to cross, and in what sequence they decide to conduct experiments. Moreover, because players are equipped with different tools that they can use to find out phenotypes and genotypes and breed organisms at their own choices, it provides an opportunity to shift between the different levels of genetics knowledge to solve problems in *Radix*. Previous research indicates that students often fail to correctly describe the nature of simple dominant and recessive patterns and can depict only monogenic inheritance (Shaw, Van Horne, Zhang, & Boughman, 2008), and mostly have

difficulties in solving dihybrid cross problems (Awang-Kanak, Masnoddin, Matawali, Daud, & Jumat, 2016). For this reason, this research attempted to delve into players' exploration of the inheritance patterns of dihybrid crosses.

The goal of this study was to identify approaches to GLA for open-ended gameplay, in order to gain insight into players' content-specific gameplay decisions and distinguish students with different approaches and varying levels of understanding. The first step was to select high-level approaches that reflect the game design elements of exploratory, open-ended games and also build on proven methods of GLA. In *Radix* and other open world games, players have agency to choose the best tools to solve a problem, and complete tasks using a variety of strategies, so the chosen approaches align with those design choices. They are also based on established GLA techniques, many of which have been highlighted as exemplars by Owen & Baker (2019). These include looking at tool use to track students' progress (Butler & Banerjee, 2014), identifying player pathways through a simulation (Institute of play, 2013), describing player strategies across different levels (Cooper et al., 2010), and comparing the activity of distinct groups of players (O'Rourke, Butler, Liu, Ballweber, & Popovic, 2013). Combining the key aspects of our learning environment with proven game data techniques, we hypothesized that the following three approaches, when applied to the *Radix* game, would illuminate and differentiate content-specific gameplay decisions. The research questions therefore align to each approach to be tested in the study:

- RQ1. How can data about actions taken or tools used to help players complete tasks be used to identify and distinguish different types of gameplay decisions?
- RQ2. How can data about play patterns or player types be used to identify and distinguish different types of gameplay decisions?

RQ3. How can data describing tool input patterns be used to identify and distinguish different types of gameplay decisions?

Materials and methods

Learning genetics in *The Radix Endeavor*

Radix is set in a rich virtual world that contains fictional but realistic ecosystems populated with diverse flora and fauna (Figure 1). Embedded in the world are simulated biological and mathematical systems which players explore to build their understanding of key concepts in science and math. It was designed as an “open-ended simulation game” or “open world game,” which are environments in which learning happens by experimenting within a system, coming to understand that system and using it for creative expression or problem-solving (Squire & Jan, 2007). In *Radix*, player actions are balanced between guided sequences of quests and the open-ended problems within those quests that they need to solve, embodying both *emergence*, a wide range of explorable possibilities based on a set of rules, and *progression*, directed challenges of increasing difficulty, drawn from Juul’s (2011) framework. Genetics is one of the content areas in *Radix*. The genetics questline consists of a sequence of tasks that gets players to explore important concepts, starting with the traits and varieties of a species, their genotypes and phenotypes, and dominant and recessive alleles; and then moving on to more complex inheritance patterns and dihybrid crosses. The game focuses on Mendelian genetics and the core mechanic of players’ genetic explorations is the ability to breed plants and animals using a simulated genetics system. Four tools that let players interact with this system are provided:

- **Trait examiner:** This tool lets players examine a plant or animal and find out what its phenotypes are. For example, when this tool is used on a myzle flower, a player may find

out that its color is red, and its height is tall. Players choose whether to use this tool in their exploration and experimentation.

- **Trait decoder:** This tool lets players find out the genotype of a plant or animal. For example, when used on a glumbug, a player may find out that it has antenna length genes LL and toxicity genes Tt. Like the trait examiner, players choose whether to use this tool in their inquiry process.
- **Breeding station:** This tool lets players choose two parent plants or animals and breed them to obtain offspring. After dragging the parents into the breeder, players can then see some basic statistics on the genotypes of the brood and interact with the offspring. Players choose which parents to use as inputs in the breeding station, but they cannot complete quests without using it in some way.
- **Punnett square:** This tool lets players fill in a Punnett square with the traits they are interested in and drag the alleles into the corresponding cells to complete the diagram. It is optionally available during the breeding process to help players make decisions.



Figure 1. A player exploring the forest biome of *Radix* and using tools to interact with the system.

Content-specific gameplay decisions. Conventional problem sets and test questions in genetics, especially at the secondary education level, tend to be straightforward and constrained to specific inheritance patterns and prescribed Punnett squares (close-ended problems), therefore providing constrained information about what students know and can do (Aznar & Orcajo, 2005). In contrast, within a genetics quest in *Radix*, game characters give players challenges with some narrative context, and players must then figure out how to go about solving the problem by selecting and breeding organisms. Taking monohybrid crosses as an example, breeding organisms requires two parents and each parent might be homozygous recessive (rr), heterozygous (Rr), or homozygous dominant (RR). Therefore, for each trait there can be six possible parental genotype combinations that players may choose to use in their cross (Table 1). Two genotypes (RR and Rr) contribute to a dominant trait and one genotype (rr) contributes to a

recessive trait. In this case five possible parental combinations (crosses 1 to 5) can produce offspring with a dominant trait, and three possible combinations (crosses 4 to 6) produce offspring with a recessive trait. Players can even choose to breed many generations to obtain the offspring with desired traits. If players don't know which variety of a trait is dominant, or which inheritance pattern that trait adheres to, with the use of equipped tools they can conduct experiments and make inferences based on the outcomes to those experiments. In other words, there are many possible approaches for solving an open-ended problem within a *Radix* genetics quest. We therefore define content-specific gameplay decisions in this study as ways of selecting tools and using them on different game objects to solve open-ended genetics problems.

Table 1

Possible parental combinations for monohybrid cross.

Cross	Parental combination	Offspring
1.	RR x RR	RR
2.	RR x Rr	$\frac{1}{2}$ RR, $\frac{1}{2}$ Rr
3.	RR x rr	Rr
4.	Rr x Rr	$\frac{1}{4}$ RR, $\frac{1}{2}$ Rr, $\frac{1}{4}$ rr
5.	Rr x rr	$\frac{1}{2}$ Rr, $\frac{1}{2}$ rr
6.	rr x rr	rr

“Mold in the Mines” quest in *Radix*

A detailed discussion about the design of *Radix* and its classroom implementation can be found in Klopfer et al. (2018) and Rosenheck et al. (2016). This study focuses on the 12th quest in the genetics series, entitled “Mold in the Mines,” in which a character named Parslicus asks players to figure out how to predictably breed milkflies that have both tastebuds on their feet and pointed probosces, and bring him two of those flies along with a Punnett square to show their

method. To do this, players need to catch milkflies and examine their traits, then determine which ones to breed. They choose whether to observe genotypes using the trait decoder tool, and/or phenotypes using the trait examiner tool (Figure 2). Either way they must figure out which genotypes code for which phenotypes, and which milkflies they should breed to result in the desired offspring. Tool choice is closely tied to the strategy a player elects to use, and players can find success in various ways as long as they develop a conceptual understanding of how each trait is inherited before creating the Punnett square and coming up with the required milkflies. The flow of quest tasks for “Mold in the Mines” is shown in Figure 3.

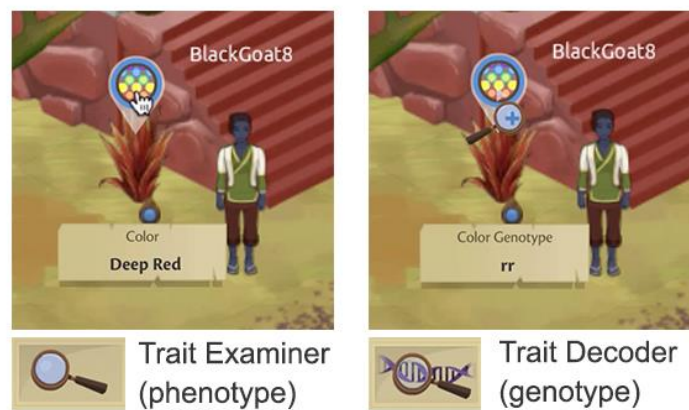


Figure 2. A player using the trait examiner (left) and trait decoder (right).

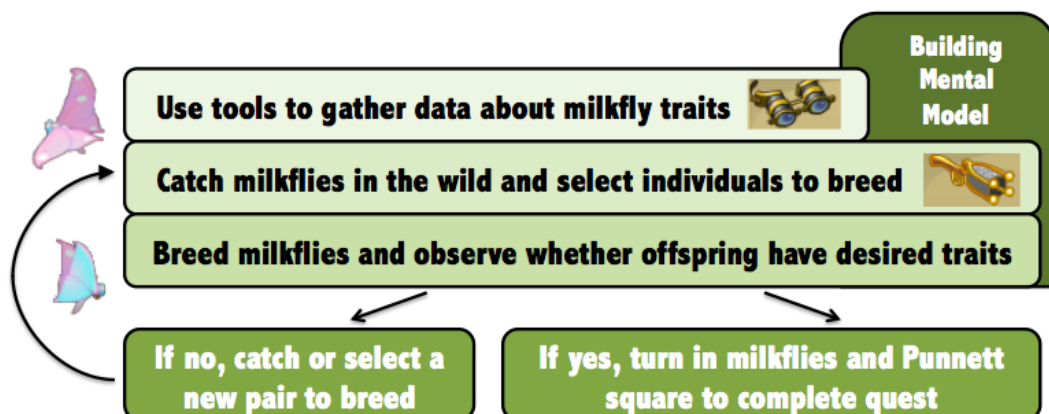


Figure 3. The flow of quest tasks for “*Mold in the Mines.*”

In order to make sense of the gameplay decisions and sequences of genetic explorations that we care about, it is important to lay out the specific game-related insights that players grapple with. Milkflies have two salient traits: presence of foot tastebuds, and proboscis shape. Which varieties are dominant and recessive is not immediately obvious and must be ascertained through experimentation. However, if a player successfully turns in flies with the desired traits, does it confidently indicate that they have explored all possible outcomes and built a robust understanding of complete dominant traits? Or might they have completed the task by understanding one part of the system—perhaps the most obvious solution, or the most sure-fire—without having fully explored the possibility space?

Content-specific gameplay decisions for “Mold in the Mines” quest. As mentioned previously, there can be six possible parental genotype combinations for each trait. Since we care about two traits in this quest, there are a total of 36 possible combinations that could be bred, and this dihybrid cross table is shown in Figure 4. In this diagram, the correct combinations, that result in at least some offspring with both desired traits, are indicated with a ✓, and the incorrect combinations are marked with a ✘. Because players can use different tools to obtain phenotype and genotype information, we take tool use into account. In some of these combinations, if a player knows the phenotypes of the parents, then the phenotypes of the offspring can be accurately predicted. For example as the presence of taste buds is a dominant trait, crossing two homozygous recessive parents (bb, bb) will definitely result in recessive offspring, without taste buds, and players can predict this by knowing that the parents had the recessive phenotype, even without having genotype information. The other five genotype combinations (any cross involving at least one BB or Bb) all have a chance of producing some offspring with taste buds,

the desired trait, and this too can be predicted if you know that at least one parent has the dominant trait. So we call all of these examiner identifiable crosses. On the other hand, a pointed proboscis is a recessive trait, so while players know that two pointed parents will produce pointed offspring, they can't know what offspring parents with round probosces will produce. This is because the outcome of a cross with a round proboscis depends on whether that parent was actually homozygous dominant (PP) or heterozygous (Pp), which can only be determined by using the trait decoder tool. We call these crosses examiner unidentifiable, because players cannot be sure of the outcome if they have only used the examiner. Looking again at Figure 4, the examiner identifiable crosses have been outlined in green, and the examiner unidentifiable crosses are outlined in red. By analyzing players' choices of parental combinations based on this dihybrid cross table using log files, we aimed to understand players' tool input patterns and broader approaches to the problem, including whether players were able to shift between the different macro/micro thought levels in order to paint a more complete picture of what they know and can do.

		Taste buds (Dominant desired)					
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB
Proboscis shape (Recessive desired)	pp-pp	✗	✓	✓	✓	✓	✓
	Pp-pp	✗	✓	✓	✓	✓	✓
	PP-pp	✗	✗	✗	✗	✗	✗
	Pp-Pp	✗	✓	✓	✓	✓	✓
	PP-Pp	✗	✗	✗	✗	✗	✗
	PP-PP	✗	✗	✗	✗	✗	✗
	PP-PP	✗	✗	✗	✗	✗	✗

Examiner Identifiable

Examiner Unidentifiable

✓ : Correct parental genotype

✗ : Incorrect parental genotype

Figure 4. Dihybrid cross table for the two relevant traits, showing correct and incorrect, and examiner identifiable and examiner unidentifiable genotype combinations.

Data collection

The game collected back-end data for all users as they played, storing data securely in a database hosted on an Amazon cloud server. This study focuses on the player usage data, in particular tool usage and quest attempts. In terms of tool usage, an event was logged every time a user used the trait examiner, trait decoder, breeding station, or Punnett square functions in the game. An event was also logged every time a player attempted to turn in a quest (hand in required objects to complete a quest). For all of these events, the inputs to the tools, that is the plants or animals the player chose to manipulate and their attributes, were logged as well. This enabled us to conduct more quest-specific and content-specific analysis probing players' understanding and strategies. Finally, each quest attempt was logged and marked as a success or failure.

All events logged are associated with a unique user id and a timestamp. Using this information each player's tool usage and quest attempts could be sequenced to examine not only their action frequencies but their quest progression as well. For example, the data can show how many times a player submits the wrong type of object, and how many times they submit the wrong variety, before ultimately succeeding. In addition, the data can also show which tools players use, how often, and in what order. All of this information has enabled researchers to find play patterns across all users who worked on this genetics quest.

Participants

The *Radix* project included a pilot that ran from January 2014 through August 2015. During

this time high school math and biology teachers were invited to use the game as part of their curriculum, implementing it in whatever way they felt best fit their students' needs. This meant some required students to play a certain amount of time or certain number of quests, while others let students explore any parts of the world that interested them as more of an enrichment activity. Across these contexts, this study includes data from 238 participants who undertook the "Mold in the Mines" quest. We chose this particular quest for analysis because it challenges players to think through a dihybrid cross, breeding for two desired traits, which creates a problem space with a wider variety of possible approaches and strategies.

Data analysis

Unless stated otherwise, student-level data, including the cumulative tool use up to the last quest event, genotypes/phenotypes of milkflies identified by each tool, and quest completion status, were used throughout the following analyses. Event-level data, collected at intermediate points throughout the data collection period, were utilized as appropriate. Descriptive statistics and visual illustration were provided to summarize the amount of tool use stratified by students' quest completion status. The association between quest completion, a binary outcome, and tool usage, an ordinal outcome, was evaluated using polychoric correlation (Ekström, 2011). A Gaussian mixture model (Hastie, Tibshirani, & Friedman, 2009) was utilized to cluster students into homogeneous subgroups based on the amount of tool use. The number of clusters was chosen using a likelihood ratio test. Upon completion of clustering, all subsequent analyses were conducted stratified by subgroup. Frequency distributions, including the number and percentage of milkfly genotypes identified by various tools, were summarized. Contingency tables were utilized to investigate patterns of tool use and milkfly genotypes/phenotypes. All statistical analyses were conducted in R version 3.3.0 or later (R Core Team, 2013). Two R add-on

packages polycor (Version 0.7-9) and mxitools (Version 1.1-0) were used to calculate polychromic correlations and optimize the clustering algorithm, respectively.

The data and analyzing codes that support the findings of this study are openly available in Figshare (Cheng, 2020).

Results

Tool use for all participants

The first research question and analytic approach focuses on actions that help players complete tasks, which in the context of *Radix* is operationalized as their choice of tools and amount of tool use, and whether those actions contributed to success on the quest. To implement this, we began by splitting participants into groups based on whether they completed the quest. In the “completers” group, there were 212 (89%) who completed the quest eventually, with 102 (102/212=48.1%) of them without any failed attempts. Next we looked at the cumulative number of uses of each tool by quest completion category, to understand more about what players in each group were doing. Figure 5 shows that for completers, the tool that was used most often was the trait examiner (mean=7.7), which was used more than twice as much as the trait decoder (mean=3.0). This suggests that the trait examiner is an important tool for completing the quest. In contrast, non-completers used the examiner (mean=3.8) and decoder (mean=4.0) almost the same amount, showing that players’ choice of tools could be related to their quest outcomes. We further examined the association between successful turn-ins and the number of uses of each tool (Table 2). The results show significantly positive correlations between quest completion and greater use of the examiner ($r^*=0.56$, $p<0.01$) and breeding station ($r^*=0.46$, $p<0.01$), while the use of the decoder ($r^*=-0.09$, $p>0.05$) and Punnett Square ($r^*=0.03$, $p>0.05$) were only slightly correlated with successful turn-ins. The game requires players to use the breeding station in order

to complete the quest because that is the only way to breed milkflies, so the more salient result here is the fact that the examiner, and not the decoder, appears to be the major tool choice for completing this quest—even though the decoder provides more detailed genotype information that could be useful to players.

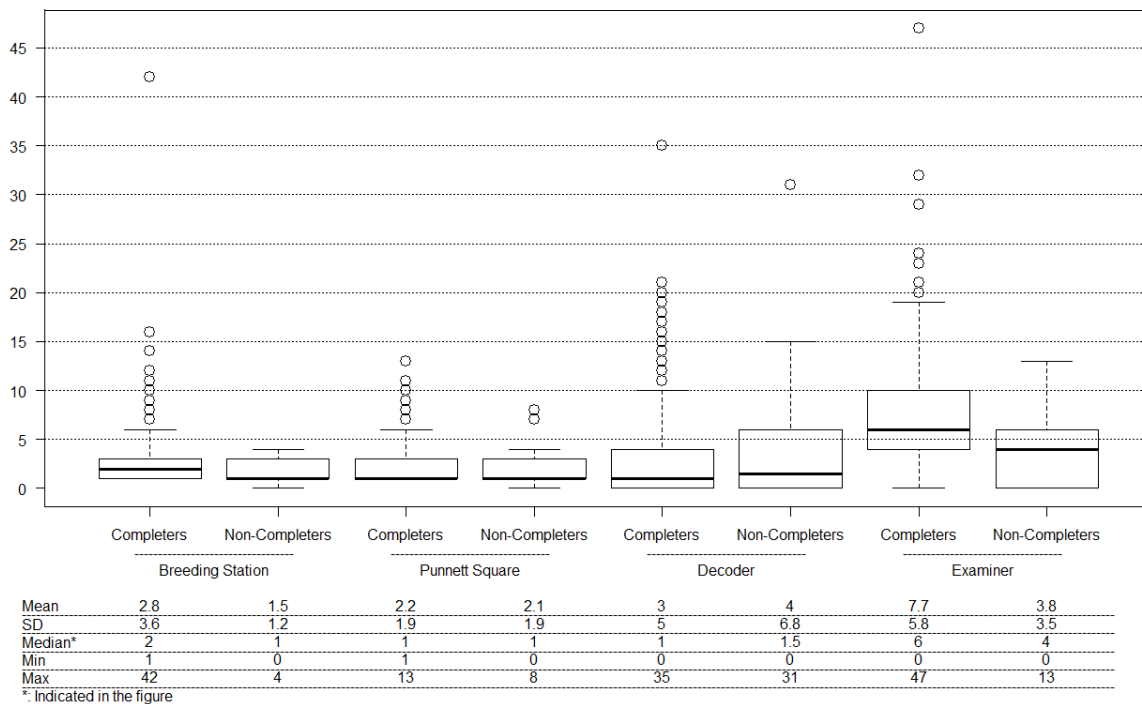


Figure 5. Cumulative number of uses of each tool by quest completion category.

Table 2.

The correlation coefficients between successful turn-ins and the number of uses of each tool (n=212).

	Examiner use	Decoder use	Punnett tool use	Breeding station use
Correlation (r*)	0.56**	-0.09	0.03	0.46**
95% CI	(0.36, 0.76)	(-0.28, 0.11)	(-0.19, 0.25)	(0.19, 0.73)

**p<0.01

Separation into discrete player types

The second research question addresses play patterns, which we identified as player types distinguished by completion status combined with amount of tool use. In order to unpack the play patterns of the completers, we used the two-dimensional scatter diagrams that compare the level of usage across tools (Figure 6). Here we found that completers actually seem to fall into two subgroups: the first group, in the solid circles, used tools fewer times in total, whereas the second group, in the dotted circles, used tools a greater number of times in order to succeed. Based on this general pattern, we then followed a more formal procedure to identify the two types of completers. Gaussian mixture model was used as the analytic method to select the number of subgroups as well as identify the membership of each completer and assign them into the most likely subgroup. As anticipated, two subgroups resulted, with “low tool use completers” (N=117) generally using fewer tools and “high tool use completers” (N=95) using more tools (Table 3). In addition, we see that low tool use completers experienced fewer failures (mean=0.44, SD=0.64) than high tool use completers (mean=2.52, SD=3.25). These results imply that low tool use completers might be those who fully understood the utility of each tool and had a better understanding of the genetics concepts, whereas high tool use completers were still trying to figure out how best to use the tools and did not have a full understanding of genetics. This could explain why the high tool use completers repeatedly used the tools and used more attempts before succeeding.

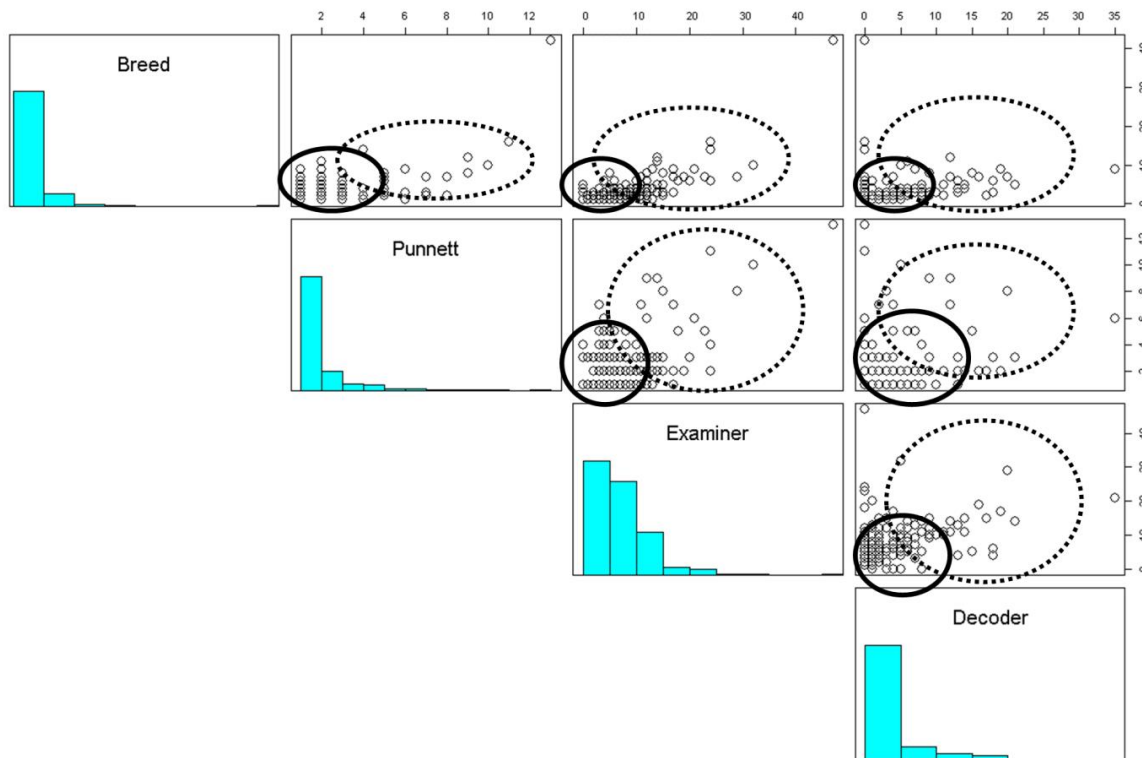


Figure 6. Comparison of level of usage across tools among quest completers. (Solid circles show low tool use completers and dotted circles show high tool use completers.)

Table 3.

Descriptive statistics of mean number of failures and different tool uses by each group.

	Low tool use completers (N=117)		High tool use completers (N=95)		Non-completers (N=26)	
	Mean	SD	Mean	SD	Mean	SD
Failures	0.44	0.64	2.52	3.25	1.46	2.20
Breeding station use	1.29	0.51	4.61	4.82	1.54	1.21
Punnett square use	1.21	0.41	3.43	2.35	2.11	1.86
Examiner use	5.69	2.76	10.18	7.49	3.85	3.51
Decoder use	1.04	1.55	5.34	6.47	3.96	6.78

Consistent with tool use results described above, both low and high tool use completers used the examiner more than the decoder. Within the low tool use completers, the examiner was used five times as much as the decoder. In addition, when we looked more deeply into the tools players used after their first failed attempt, we again found that both completer groups tended to use the examiner more than other tools, with mean uses of 1.92 and 1.94 (Table 4). This is meaningful because a failed attempt often provides players with feedback or insights into what they need to do to correct their errors, and we see here that completers still choose to use the examiner more than the decoder. On the other hand, after the first failed attempt, non-completers continued to be more likely to use the decoder than the examiner (mean=1.75 and 1.13 respectively). This behavior pattern could well be related to the reasons that non-completers never figure out their solutions to the quest. When taken together, this set of results implies that a player's choice to use the examiner or the decoder, and therefore whether to work with phenotypes or genotypes, is related to their ability to solve genetics problems.

Table 4.

The mean number of additional tool uses among participants after the first failed attempt.

	Low Tool Use Completers (N=117)	High Tool Use Completers (N=95)	Non-completers (N=26)
Breeding station use	0.72	0.77	0.34
Punnett square use	0.36	0.69	0.61
Examiner use	1.92	1.94	1.13
Decoder use	0.32	0.77	1.75

Distinguishing players' tool input patterns

The third research question looks at players' various tool input patterns, which in this quest meant the choices about which individual organisms they used with which tools, and for which

types of problems. In some ways, it is surprising that the quest can be completed more easily by using the trait examiner instead of the trait decoder, given that the decoder provides more detailed information about a creature's genotype, and the examiner only provides its phenotype. To understand why these players' approaches involved choosing to use the examiner so often and to confirm whether the quest can be solved by mainly using the trait examiner, we looked further into their interactions with the genetics content—namely which combinations of genotypes they chose to use for breeding and creating Punnett squares. These decisions are at the heart of characterizing tool input patterns, and identifying content-specific gameplay decisions.

From Figure 4 we understand the possibility space for genetic crosses in this quest, and now we dig into the genotype combinations that players chose for breeding and creating Punnett squares, and differences between low tool use completers, high tool use completers, and non-completers. In Figure 7 we can see that for both low and high tool use completers, almost three fourths of crosses that students chose for breeding were both examiner identifiable and correct (73.5% and 72.6% respectively). But for non-completers, more than 80% of crosses chosen were examiner unidentifiable, and 50% were incorrect. The crosses players used for creating Punnett squares, shown in Figure 8, reveal a very similar pattern. 75.2% of crosses chosen by low tool use completers, and 81.8% chosen by high tool use completers, were examiner identifiable. But the opposite is true for non-completers, where 85% of crosses were examiner unidentifiable, and 50% were incorrect. So we see that completers find success with the examiner not necessarily because it's the most informative tool, but because they know that in certain situations it gives them just enough information. Their choice seems to revolve around optimizing the information they have and using it efficiently, rather than trying to uncover all the information possible. These illuminating insights into player behavior show that examining the content-specific

decisions made by players reveals a more complete picture than analyzing their tool use alone. Knowing that a student completed the quest indicates a certain level of genetics knowledge, and knowing which tools they used describes their gameplay. To go deeper, knowing exactly how they used those tools and for what purposes can more fully explain their content-specific gameplay decisions, which tells us much more about their understanding of genetics as a system, and about their ability to solve complex problems.

Author accepted manuscript

Breeding Station: Low Tool Use Completers

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	3 ✓	5 ✓	22 ✓	33 ✓	23 ✓	= 86 (73.5%)	
	Pp-pp	3 ✓	2 ✓	5 ✓	10 ✓	9 ✓	= 29 (24.8%)	
	PP-pp							
	Pp-Pp	✓	✓	1 ✓	1 ✓	✓	= 2 (1.7%)	
	PP-Pp							
	PP-PP							
Total = 117								

Breeding Station: High Tool Use Completers

Examiner Identifiable
Examiner Unidentifiable
 ✓ : Correct parental genotype

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	1 ✓	4 ✓	14 ✓	26 ✓	24 ✓	= 69 (72.6%)	
	Pp-pp	3 ✓	4 ✓	2 ✓	6 ✓	5 ✓	= 20 (21.1%)	
	PP-pp							
	Pp-Pp	✓	✓	2 ✓	2 ✓	2 ✓	= 6 (6.3%)	
	PP-Pp							
	PP-PP							
Total = 95								

Breeding Station: Non-completers

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	2	✓	✓	1 ✓	✓	✓	= 3 (15.0%)
	Pp-pp	1 ✓	1 ✓	2 ✓	2 ✓	✓		= 6 (30.0%)
	PP-pp	1				1		= 2 (10.0%)
	Pp-Pp	✓	✓	✓	✓	1 ✓		= 1 (5.0%)
	PP-Pp	1			1	1		= 3 (15.0%)
	PP-PP				1	1	3	= 5 (25.0%)
Total = 20								

Figure 7. Cross tables showing the genotype combinations for breeding that are examiner identifiable and unidentifiable.

Punnett Tool: Low Tool Use Completers

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	2 ✓	4 ✓	23 ✓	33 ✓	26 ✓	= 88 (75.2%)	
	Pp-pp	4 ✓	2 ✓	3 ✓	9 ✓	9 ✓	= 27 (23.1%)	
	PP-pp							
	Pp-Pp	✓	✓	1 ✓	1 ✓	✓	= 2 (1.7%)	
	PP-Pp							
	PP-PP							
Total = 117								

Punnett Tool: High Tool Use Completers

Examiner
Identifiable

Examiner
Unidentifiable

✓ : Correct parental genotype

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	✓	3 ✓	20 ✓	20 ✓	34 ✓	= 77 (81.8%)	
	Pp-pp	1 ✓	3 ✓	2 ✓	6 ✓	1 ✓	= 13 (13.7%)	
	PP-pp							
	Pp-Pp	✓	✓	1 ✓	3 ✓	1 ✓	= 5 (5.3%)	
	PP-Pp							
	PP-PP							
Total = 95								

Punnett Tool: Non-completers

		Taste buds (Dominant desired)						
		bb-bb	Bb-bb	BB-bb	Bb-Bb	BB-Bb	BB-BB	
Proboscis shape (Recessive desired)	pp-pp	2	✓	✓	1 ✓	✓	✓	= 3 (15.0%)
	Pp-pp	1 ✓	2 ✓	1 ✓	2 ✓	✓	✓	= 6 (30.0%)
	PP-pp	1				1		= 2 (10.0%)
	Pp-Pp	✓	✓	✓	✓	✓	1 ✓	= 1 (5.0%)
	PP-Pp	1			1	1		= 3 (15.0%)
	PP-PP	1					4	= 5 (25.0%)
Total = 20								

Figure 8. Cross tables showing the genotype combinations for Punnett tool use that are examiner identifiable and unidentifiable.

Discussion

The analysis presented here shows that all three approaches did in fact contribute to

identifying gameplay decisions related to genetics, and helped to distinguish students using different approaches to complete game tasks. In particular, the examination of tool input patterns, which also incorporated tool use and player types, illuminated the most meaningful distinctions between students making different types of decisions.

In response to RQ1 and RQ2, the data on tool use and player types show that these approaches can be useful in presenting a picture of different players' experiences, by revealing the three qualitatively different player groups described above. The low tool use completers are players that demonstrate an efficient strategy—they know what they're trying to do, and it doesn't take them very many tool uses or quest attempts to accomplish it. High tool use completers may not know right away how to solve a problem, but they have some idea of a strategy and they are willing to try things and figure it out. They go through more failed attempts and need to use the tools more times before they get it right, but they are willing and able to go through that problem-solving process, demonstrating another effective strategy with a few more steps in the thought process. The non-completers on the other hand, seem to struggle in two ways. They likely lack an understanding of the genetic system, and perhaps also the tools available to them. But more significantly they don't seem to try as many times or as many different ways as the high tool use completers. They may not have the problem-solving mindset that would make them comfortable experimenting and figuring things out on their own.

With the identification of these patterns and a method to classify them based on player actions, real-time analytics could be built into *Radix* or other serious games to classify each player's style or strategy. Designers could use this information to tailor the game's feedback and scaffolding to each player, for example allowing high tool use completers to engage in productive struggle on subsequent quests, while letting the quest-giver give more guidance or

feedback to non-completers at risk of giving up. Or they could revise the design of the quest, constraining the task further to make it impossible to complete without utilizing genotype information. In this case the quest might go further than requiring a trait, instead asking specifically for milkflies that are heterozygous for one or both traits, or asking for a whole brood of offspring with certain distributions of traits. Equally important to these design decisions, the classifications could be shared with teachers. Combined with what they already know about students, this additional level of data would prepare teachers to address individual students' needs in terms of genetics knowledge and problem-solving strategies, in ways best suited to each type of learner. Further, knowing which crosses students had conducted and which were left unexplored would mean that a teacher could challenge them to explore those in other ways both in and outside of the game. One of the unique aspects of *Radix* is that it is an inquiry-based game that presents opportunities to explore multiple solutions to a task, and for this reason being able to identify whether players are struggling with that exploration process, and what their process looks like, is essential. Because of the open-ended nature of this learning environment, evidence of students' varying decisions can be captured and differentiated, and in answer to RQ3 we see that it can be classified into distinct tool input patterns that tell us about players' approaches to the task and understanding of genetic inheritance.

For the low tool use completers, the group that seems to have a good grasp of genetic inheritance patterns, we not only want to know that they can use genetics to solve problems, but how they are thinking about those problems. The learning objective for the "Mold in the Mines" quest was to understand how genotypes in dominant and recessive traits combine to result in each phenotype. This quest involved a dihybrid task, with two desired traits, which we expected would necessitate more use of the trait decoder. However, players didn't approach the task this

way—instead it seems that players were not even looking at the milkflies' genotypes. Rather than setting a goal of gathering genetic information about the milkflies and then choosing the appropriate tool for the job (the decoder), players seem to choose the tool they want to use (the examiner) and then figure out how to accomplish the task within the constraints of what that tool can do. While the data logs don't tell us why players make these choices, we can guess that it may be due to a perception of the examiner as being easier to use, compared with digging into genotypes which seem more complex. Phenotypes are visible and therefore more concrete, whereas genotypes are more abstract and require the manipulation of symbols (Johnstone, 1991). The phenomena involving multi-level thinking are usually confusing to students (Wilensky & Resnick, 1999), and it seems that in *Radix* players did not shift between the different macro/micro thought levels. Using the simpler tool, and identifying which are the sure-fire crosses—the ones that are examiner identifiable and can avoid seemingly unpredictable heterozygotes—may be preferable to dealing with genotypes and the decoder. While this is not the best method in terms of genetics learning, it is valid given the framing and goals of the game.

Using a combination of the three approaches we hypothesized would be illuminating, our analysis has revealed learners' tendency to choose sure-fire crosses that are simpler to understand. This is an example of how players will adapt a learning environment to their own goals and purposes, a pattern that is more difficult to ascertain with small-scale observations than with large-scale in situ data analysis. While the task does provide an environment to explore all possible crosses, the motivation is only there to explore the most sure-fire, examiner identifiable crosses. So for players that choose those crosses right away and breed them correctly, we can feel confident that they understand that aspect, for example homozygous recessive crosses. But, we can't assume anything about their understanding of what offspring would result from

heterozygous parents. In a more traditional paper-based problem set, each concept would be targeted individually so knowing which crosses a student worked with would be a given. However, real-world scientists do not have such specific instructions and student-centered pedagogy greatly values the cognitive task of figuring out which experiments to do and how to set them up. While it may not be surprising that students will use the simplest tool possible to solve a problem, this content-specific approach to analysis illuminates who is thinking in which way, so that students can reflect on their choices and teachers can further challenge them. While much is already understood about students' challenges in thinking about genetics, these approaches could likely be used to distinguish types of decisions and interaction patterns in biology or various other domains. Making sense of complicated sequences of events is an affordance of GLA that can strengthen how people learn using digital experiences.

Taking the "Mold in the Mines" data analysis as a case study, we can see that GLA can help us describe inquiry processes that don't always occur in more traditional pedagogies. The ability to understand students' choices around which tools to use and what experiments to run means that teachers and designers have more information with which to support student-centered pedagogies such as inquiry learning. From this example of game design and data analysis, it is clear that game data can tell us a great deal—but it needs to be content-specific and granular enough to get at what students are doing in relation to the key concepts involved. Completion data and tool use could easily be analyzed across all quests in the game, but it wouldn't help us understand how students go about exploring genetics in an inquiry environment—that they have a tendency to work with phenotypes over genotypes, even when it doesn't provide as much detail about a problem. With the use of more generalized metrics, for example simple quest completion, many of the distinctions identified in this study would go unnoticed. Data collection

at the level of content-specific decisions is more challenging to design and time-consuming to analyze because there is no one size fits all formula, but this is where the rich evidence for learning lies. Too often we see game analytics that report on completion, perhaps number of attempts and time spent, and infer from that what the players have learned or can do. This is only the first step in embedding assessment into games, but is not enough to provide evidence of content-specific skills and processes. As we have seen, more detailed data on player decisions, combined with general play patterns, can provide insights into players' problem-solving approaches. This could be further developed into a robust way to assess inquiry and problem-solving skills and other related science practices.

The work discussed in this paper is meant to be a start to identifying types of analyses that can illuminate content-specific gameplay decisions in a wider variety of serious games. In this study we used three key methods. The first was to identify actions that help players complete tasks, thereby painting a picture of what successful players are doing that others aren't. The second was to classify play styles—based on actions, progress, or other factors. And the third was to examine tool input patterns that players demonstrated, paying special attention to the inputs certain groups did not use, thereby revealing the concepts they did not fully explore and strategies they did not utilize, which in turn presents opportunities for deepening their understanding. While these types of analyses must be tailored to any individual game and its tasks, and specific statistical methods must be selected, this study has shown their utility in the context of one game. We therefore present them as guiding questions for designing analysis of open-ended game experiences, while recognizing the need for future work to expand on these approaches and test their application to other learning games.

Illuminating students' content-specific gameplay decisions requires a context that is open-ended enough for students to make choices that demonstrate their thinking. Doing this also requires analysis methods that are granular enough to look at the specific choices across player types most relevant to that context. This paper has shown how the context of *Radix* and analysis of choices in genetic crosses can make visible how students think about using phenotype and genotype information to perform dihybrid crosses. With the development of more open-ended, inquiry-based games and interactive learning environments, along with the establishment of guiding questions for analysis, rich patterns in solving content-specific problems can be brought to light, thereby deepening the reflection and learning in technology-rich environments.

Acknowledgement

This research was funded by the Bill and Melinda Gates Foundation and the Ministry of Science and Technology (MOST), Taiwan, under grant contract no. 104-2918-I-018-008 and 105-2511-S-018-015-MY5. We would like to thank the members of the *Radix* team at MIT who made this game and research possible, including Angie Tung, Susannah Gordon-Messer, and Jody Clarke-Midura.

Compliance with Ethical Standards

Conflict of Interest

The authors declare that they have no conflict of interest.

Ethics approval

This study was carried out with institutional review board (IRB) approval and in accordance with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

Informed consent

All participants were well informed about the research objectives and contents and were assured about the privacy. None of the personal identifiers was reported in the study.

References

- Alonso-Fernández, C., Calvo, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2017). Systematizing game learning analytics for serious games. Paper presented at the - 2017 *IEEE Global Engineering Education Conference (EDUCON)*, 1111-1118. doi:10.1109/EDUCON.2017.7942988
- Alonso-Fernández, C., Calvo-Morata, A., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2019). Applications of data science to game learning analytics data: A systematic literature review. *Computers & Education*, *141*, 103612. doi:<https://doi.org/10.1016/j.compedu.2019.103612>
- Awang-Kanak, F., Masnoddin, M., Matawali, A., Daud, M. A., & Jumat, N. R. (2016). Difficulties experience by science foundation students on basic mendelian genetics topic: A preliminary study. *Transactions on Science and Technology*, *3*(1-2), 283-290.
- Aznar, M. M., & Orcajo, T. I. (2005). Solving problems in genetics. *International Journal of Science Education*, *27*(1), 101-121. doi:10.1080/09500690410001673801
- Bado, N. (2019). Game-based learning pedagogy: A review of the literature. *Interactive Learning Environments*, , 1-13. doi:10.1080/10494820.2019.1683587
- Bahar, M., Johnstone, A. H., & Hansell, M. H. (1999). Revisiting learning difficulties in biology. *Journal of Biological Education*, *33*(2), 84-86. doi:10.1080/00219266.1999.9655648
- Butler, E., & Banerjee, R. (2014). Visualizing progressions for education and game design (white paper). Retrieved from https://www.ericbutler.net/assets/papers/unpub2014_visualization.pdf

- Chaudy, Y., Connolly, T., & Hainey, T. (2014). Learning analytics in serious games: A review of the literature. Paper presented at the *ECAET 2014: European Conference in the Applications of Enabling Technologies*,
- Cheng, M.-T. (2020): The radix endeavor dataset. figshare. Dataset.
<https://doi.org/10.6084/m9.figshare.12442469.v1>
- Chu, Y., & Reid, N. (2012). Genetics at school level: Addressing the difficulties. *Research in Science & Technological Education*, 30(3), 285-309. doi:10.1080/02635143.2012.732059
- Collins, A., & Stewart, J. H. (1989). The knowledge structure of mendelian genetics. *The American Biology Teacher*, 51(3), 143-149.
- Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., . . . players, F. (2010). Predicting protein structures with a multiplayer online game. *Nature*, 466(7307), 756-760. doi:10.1038/nature09304
- Duncan, R. G., & Reiser, B. J. (2007). Reasoning across ontologically distinct levels: Students' understandings of molecular genetics. *Journal of Research in Science Teaching*, 44(7), 938-959. doi:10.1002/tea.20186
- Etobro, A. B., & Banjoko, S. O. (2017). Misconceptions of genetics concepts among pre-service teachers. *Global Journal of Educational Research*, 16(2), 121-128.
- Freire, M., Serrano-Laguna, Á, Manero, B., Martínez-Ortiz, I, Moreno-Ger, P., & Fernández-Manjón, B. (2016). Game learning analytics: Learning analytics for serious games. *Learning, design, and technology* (pp. 1-29). AG, Switzerland: Springer Nature.
- Institute of play. (2013). Playtime online 21: Digging into data with SimCityEDU. Retrieved from https://www.youtube.com/watch?v=Y5_7Y7wKE6A&feature=share
- Johnstone, A. H. (1991). Why is science difficult to learn? things are seldom what they seem. *Journal of Computer Assisted Learning*, 7(2), 75-83. doi:10.1111/j.1365-2729.1991.tb00230.x

- Juul, J. (2011). *Half-real: Video games between real rules and fictional worlds* MIT press.
- Karagoz, M., & Cakir, M. (2011). Problem solving in genetics: Conceptual and procedural difficulties. *Educational Sciences: Theory and Practice*, 11(3), 1668-1674.
- Ke, F. (2016). Designing and integrating purposeful learning in game play: A systematic review. *Educational Technology Research and Development*, 64(2), 219-244. doi:10.1007/s11423-015-9418-1
- Ketelhut, D. J. (2007). The impact of student self-efficacy on scientific inquiry skills: An exploratory investigation in river city, a multi-user virtual environment. *Journal of Science Education and Technology*, 16(1), 99-111. doi:10.1007/s10956-006-9038-y
- Klopfer, E., Haas, J., Osterweil, S., & Rosenheck, L. (2018). *Resonant games*. Cambridge, MA: MIT Press.
- Knippels, M. P. J., & Waarlo, A. J. (2018). Development, uptake, and wider applicability of the yo-yo strategy in biology education research: A reappraisal. *Education Sciences*, 8(3), 129.
- Knippels, M. P. J., Waarlo, A. J., & Boersma, K. T. (2005). Design criteria for learning and teaching genetics. *Journal of Biological Education*, 39(3), 108-112. doi:10.1080/00219266.2005.9655976
- Lewis, J., & Kattmann, U. (2004). Traits, genes, particles and information: Re-visiting students' understandings of genetics. *International Journal of Science Education*, 26(2), 195-206. doi:10.1080/0950069032000072782
- Liu, M., Kang, J., Liu, S., Zou, W., & Hodson, J. (2017). Learning analytics as an assessment tool in serious games: A review of literature. In M. Ma, & A. Oikonomou (Eds.), *Serious games and edutainment applications* (pp. 537-563) Springer.
- Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015a). Serious games analytics: Theoretical framework. In C. S. Loh, Y. Sheng & D. Ifenthaler (Eds.), *Serious game analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 3-29). Switzerland: Springer.

Loh, C. S., Sheng, Y., & Ifenthaler, D. (2015b). *Serious games analytics: Methodologies for performance measurement, assessment, and improvement*. Switzerland: Springer.

NHGRI. (n.d.). Phenotype. Retrieved from <https://www.genome.gov/genetics-glossary/Phenotype>

Orcajo, T. I., & Aznar, M. M. (2005). Solving problems in genetics II: Conceptual restructuring. *International Journal of Science Education*, 27(12), 1495-1519.
doi:10.1080/09500690500186584

O'Rourke, E., Butler, E., Liu, Y., Ballweber, C., & Popovic, Z. (2013). The effects of age on player behavior in educational games. Paper presented at the *The 8th International Conference on the Foundations of Digital Games*, 158-165.

Owen, V. E., & Baker, R. S. (2019). Learning analytics for games. In J. L. Plass, R. E. Mayer & B. D. Homer (Eds.), *Handbook of game-based learning* (pp. 513-536). Cambridge, MA: MIT.

Qian, M., & Clark, K. R. (2016). Game-based learning and 21st century skills: A review of recent research. *Computers in Human Behavior*, 63, 50-58.
doi:<https://doi.org/10.1016/j.chb.2016.05.023>

Rosenheck, L., Gordon-Messer, S., Clarke-Midura, J., & Klopfer, E. (2016). Design and implementation of an MMO: Approaches to support inquiry learning with games. In D. Russell, & J. M. Laffey (Eds.), *Handbook of research on gaming trends in P-12 education* (pp. 33-54). Hershey, PA: IGI Global.

Rowe, E., Asbell-Clarke, J., & Baker, R. S. (2015). Serious games analytics to measure implicit science learning. In C. S. Loh, Y. Sheng & D. Ifenthaler (Eds.), *Serious games analytics: Methodologies for performance measurement, assessment, and improvement* (pp. 343-360). Cham: Springer International Publishing. doi:10.1007/978-3-319-05834-4_15 Retrieved from https://doi.org/10.1007/978-3-319-05834-4_15

- Shaw, K. R. M., Van Horne, K., Zhang, H., & Boughman, J. (2008). Essay contest reveals misconceptions of high school students in genetics content. *Genetics*, *178*(3), 1157-1168. doi:10.1534/genetics.107.084194
- Squire, K. D., & Jan, M. (2007). Mad city mystery: Developing scientific argumentation skills with a place-based augmented reality game on handheld computers. *Journal of Science Education and Technology*, *16*(1), 5-29.
- Steinkuehler, C. A., & Duncan, S. (2008). Scientific habits of mind in virtual worlds. *Journal of Science Education and Technology*, *17*(6), 530-543. doi:10.1007/s10956-008-9120-8
- Tsui, C., & Treagust, D. (2010). Evaluating secondary students' scientific reasoning in genetics using a two-tier diagnostic instrument. *International Journal of Science Education*, *32*(8), 1073-1098. doi:10.1080/09500690902951429
- Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3-19. doi:10.1023/A:1009421303064