Journal of Science Education and Technology (2021) 30:511–528
https://doi.org/10.1007/s10956-020-09896-8

Check for updates

Modeling Secondary Students' Genetics Learning in a Game-Based Environment: Integrating the Expectancy-Value Theory of Achievement Motivation and Flow Theory

Arif Rachmatullah[1] · Frieda Reichsman[2] · Trudi Lord[2] · Chad Dorsey[2] · Bradford Mott[3] · James Lester[3] · Eric Wiebe[1]

Accepted: 21 December 2020 / Published online: 21 January 2021
© The Author(s), under exclusive licence to Springer Nature B.V. part of Springer Nature 2021

Abstract
This study examined students' genetics learning in a game-based environment by exploring the connections between the expectancy-value theory of achievement motivation and flow theory. A total of 394 secondary school students were recruited and learned genetics concepts through interacting with a game-based learning environment. We measured their science self-efficacy, science outcome-expectancy beliefs, flow experience, feelings of frustration, and conceptual understanding before and after playing the game, as well as their game satisfaction. Mixed-model ANOVA, correlation tests, and path analysis were run to answer our research questions. Based on the results, we found that the game had a significant impact on students' conceptual understanding of genetics. We also found an acceptable statistical model of the integration between the two theories. Flow experience and in-game performance significantly impacted students' posttest scores. Moreover, science outcome-expectancy belief was found to be a significant predictor of students' flow experiences. In contrast, science self-efficacy and pretest scores were found to be the most significant factors influencing the feeling of frustration during the game. The results have practical implications with regard to the positive role that an adaptive game-based genetics learning environment might play in the science classroom. Findings also underscore the role the teacher should play in establishing productive outcome expectations for students prior to and during gameplay.

Keywords Outcome-expectancy · Flow theory · Game-based learning · Genetics · Self-efficacy

✉ Arif Rachmatullah
arifraach@gmail.com; arachma@ncsu.edu

Frieda Reichsman
freichsman@concord.org

Trudi Lord
tlord@concord.org

Chad Dorsey
cdorsey@concord.org

Bradford Mott
bwmott@ncsu.edu

James Lester
lester@ncsu.edu

Eric Wiebe
wiebe@ncsu.edu

1 Department of STEM Education, North Carolina State University, Raleigh, NC, USA

2 The Concord Consortium, Concord, MA, USA

3 Department of Computer Science, North Carolina State University, Raleigh, NC, USA

Introduction

Digital games are among the most popular and influential media used in science instruction across K-12 education (Riopel et al. 2020). One reason for their popularity is that digital games can dynamically create rich, interactive learning experiences around abstract science concepts, such as those in molecular genetics, allowing students to better envision and grasp them (Cheng et al. 2014). Digital games also allow students to perform authentic and complex science experiments repeatedly without concerns for any life consequences (Cheng and Annetta 2012). It follows that research has broadly argued that digital games are significant alternative tools for teaching and learning activities, particularly in science (Riopel et al. 2020). Hence, research suggested that digital games can improve not only students' scientific conceptual understanding (Riopel et al. 2020) but also their affective and motivational orientation toward science (Li and Tsai 2013; Vogel et al. 2006), as well as their scientific practices (Bressler and Bodzin 2016).

Studies of high-school students by Gericke and Walberg (2013) showed that the perception of abstractness and image are among the difficulties that students encounter when learning complex, microscopic science topics such as genetics. Rotbain et al. (2008), who conducted a study on the use of computer animation in high-school genetics, suggested overcoming such difficulties by providing students with a visualization of genetics processes, such as translation of particular genes to an organism's physical appearance. Using computer animation and interactive models to teach genetics concepts by increasing agency and providing a more dynamic simulation through digital games (e.g., Annetta et al. 2009; Marbach-Ad et al. 2008; Kim et al. 2015) has been shown promising findings. Contextualizing such genetics content and media elements within digital games have been demonstrated to significantly improve students' genetics learning, particularly their understanding of protein synthesis, sexual reproduction, and inheritance (Wilson et al. 2018).

A recent meta-analysis study by Riopel et al. (2020) involving 79 empirical studies on game-based science learning demonstrates that the use of digital games in science learning could significantly support gains in declarative and procedural knowledge and knowledge retention. The authors also found that the use of digital games was more effective in teaching certain science subjects, such as life science and physics, when implemented in shorter durations (e.g., less than a week, two sessions). One intriguing finding relevant to the present study concerns the influence of the level of user control over content on learning gains. Riopel et al. (2020) found that digital games that provided users with more control over content positively correlate to learning gains. With individualized control, students' values and goals for learning can drive the navigation of their learning which, in turn, can lead them to more enjoyable learning experiences. Aligning students' own valuation and purpose of learning with their feelings of enjoyment leads to immersion in the activities with greater intrinsic motivation, promoting a "flow experience" (Bressler and Bodzin 2016; Csikszentmihalyi 2014).

Flow experience is a well-studied aspect of game-based learning, which refers to students' optimal experience when they are deeply engaged in an activity whose challenges are congruent with their skills (Sharek and Wiebe 2014). "Flow experience" was first coined by Csikszentmihalyi (1990) in the context of his flow theory. Flow theory also articulates how the feeling of frustration resulting from challenging tasks drives individuals to stay focused on the challenge. However, ongoing frustration may gradually disrupt focus attention if one keeps experiencing it throughout gameplay (Melhárt 2018). Research has identified mixed findings around the relationship between feelings of frustration and learning gains in game-based environments—some found that the two have a negative relationship (e.g., DeFalco et al. 2018; Hone 2006) while others found that feelings of frustration do not relate to students' learning (e.g., Baker et al. 2010; Shute et al. 2015).

Additionally, some studies have identified several factors that influence individuals' flow and frustration in game-based learning, such as self-efficacy and outcome expectancy beliefs. Studies show that self-efficacy and expectation play a significant role in influencing flow experience and frustration in digital game–based science learning, especially with regard to keeping individuals on task (Burak 2014; Hung et al. 2015; Melhárt 2018). This implies that the expectancy-value theory of achievement motivation (Wigfield and Eccles, 2000) may also have a connection with flow theory and help explain student learning in a game-based environment. However, the combination of these two theories to guide the study of learning outcomes in digital games seems to be a novel approach with regard to the current state of the educational research literature.

Thus, the current study sought to examine whether these two theories can be integrated into one explanatory statistical model with the goal of better understanding students' genetics learning in a game-based environment. This proposed work would help identify the critical components of the connection between the two theories so that practitioners and researchers would be able to develop a more impactful learning intervention. The following research questions guided this current study:

1. Does a digital game–based learning environment increase students' understanding of genetics concepts?
2. Is there any association between students' game experiences with their learning gains?
3. Are there any correlations between in- and out-of-game factors?
4. How does the intercorrelation of the in- and out-of-game factors described by a path analysis explain students' genetics learning in a game-based environment?

Theoretical Framework

Expectancy-Value Theory of Achievement Motivation

The expectancy-value theory of achievement motivation (EVT) posits that "individuals' choice, persistence, and performance can be explained by their beliefs about how well they will do on the activity and the extent to which they value the activity" (Wigfield and Eccles 2000, p. 68). In their later work, Eccles and Wigfield (2002) stated that an individual actively and regularly assesses the attainment of specific goals, as well as the cost-and-benefit values of such accomplishments. However, the present study focused less on the value component of EVT, and more on the expectation of success: competence and outcome-expectancy beliefs (Eccles and Wigfield 2002).

Competence belief or self-concept is usually used interchangeably with self-efficacy. Self-efficacy is beliefs about one's ability to perform, execute, and complete a particular task (Pajares 1996). Bandura's (1997) social-cognitive theory similarly conceptualizes self-efficacy as is used in the EVT model; thus, EVT usually complements social-cognitive theory (Unfried et al. 2015). While both theories emphasize perceived confidence, Bong and Skaalvik (2003) and Pajares (1997) noted that Bandura's self-efficacy is more activity-specific, while EVT's self-efficacy is more general. We were more interested in EVT's interpretation of self-efficacy in order to examine how students' beliefs in their general science competence influence specific learning activities. Wigfield and Eccles (2000) conceptualized outcome-expectancy beliefs as learners' beliefs about what they will do in nearly immediate or more-extended future events. Self-efficacy and outcome-expectancy beliefs share a similar feature, in that both depend on one's current ability or competency.

Researchers have widely used EVT to examine students' persistence in STEM careers (e.g., Guo et al. 2017; Lauermann et al. 2017; Wiebe et al. 2018) and difficult STEM-related tasks (e.g., Abraham and Barker 2015). Similarly, studies have used self-efficacy to explain students' learning in game-based environments (e.g., Eseryel et al. 2014; Su 2016). Less common has been the use of EVT—particularly the combination of self-efficacy and outcome expectancy beliefs—in an explanatory model of student learning in a game-based environment. We believed that these two components interact with flow experience to influence students' learning in such learning environments.

Flow Theory

Csikszentmihalyi (1990) first coined the term "flow theory" to describe a phenomenon of optimal experience whereby individuals deeply engage in a fun and enjoyable task. Individuals are in a flow state when they encounter a challenging task and judge accomplishing the task as valuable. Csikszentmihalyi (1988, 2014) posited that in a flow state, people are intrinsically motivated, feel in control, maintain focused concentration, and do not monitor time. Nakamura and Csikszenmihalyi (2014) also explained that individuals seek to replicate flow experiences because they are intrinsically rewarding.

Additionally, Przybylski et al. (2010) and Sharek and Wiebe (2014) connected the idea of willingness to pursue obtainable challenges in a flow state to motivation, self-efficacy, and engagement. According to these researchers, motivation to perform a particular task leads to engagement. Once an individual has engaged with the task, the opportunity arises to enter a flow state. One would maintain a flow state as long as the challenging task is relatively equal to the individual's actual and perceived ability; otherwise, the individual would enter disengagement (O'Brien and Toms 2008), feeling frustrated if the task is too difficult to perform (i.e., a frustration state). In contrast, if the task were too easy, the individual would enter a state of boredom. Therefore, in the context of the digital game–based environment, game developers need to consider how to maintain the optimal level of challenge so that the users can stay in the flow state.

Hypothesized Model

In the present study, we have hypothesized that the integration of the two theories—EVT and flow theory—could help explain the psychological mechanisms underlying students' genetics learning in a digital game–based environment. Below are the reviews on previous studies that we used to generate a hypothesized model presented in Fig. 1.

Science Self-efficacy, Outcome-Expectancy Belief, and Cognitive Score

According to EVT (Eccles and Wigfield 2002; Wigfield and Eccles 2000), students' self-efficacy and outcome-expectancy beliefs are positive predictors of achievement. In relation to science learning, Uçar and Sungur (2017) explored seventh-grade students' science self-efficacy and their science achievement and found that higher self-efficacy contributed to a higher chance of students succeeding in science class. Su's (2016) work in a game-based learning environment showed that motivation positively influences cognitive performance. Moreover, using a structural equation modeling (SEM) technique to understand the relationship between cognitive scores before, during, and after learning in a game-based environment, Shute et al. (2015) showed that these three scores were significantly predictive of one another. Based on these studies, we predicted that both science self-efficacy and outcome expectancy would have a positive impact on students' cognitive outcomes, either before, during, or after the gameplay. Finally, the earlier cited literature on game-based science learning (e.g., Cheng et al. 2015; Riopel et al. 2020) predicts that gameplay will have a positive impact on learning outcomes. The followings are more detailed hypotheses:

- H_1. Science self-efficacy has a positive impact on the genetics pretest score.
- H_2. Science outcome-expectancy belief has a positive impact on the genetics posttest score.
- H_3. Genetics pretest score is positively predictive of the posttest score.

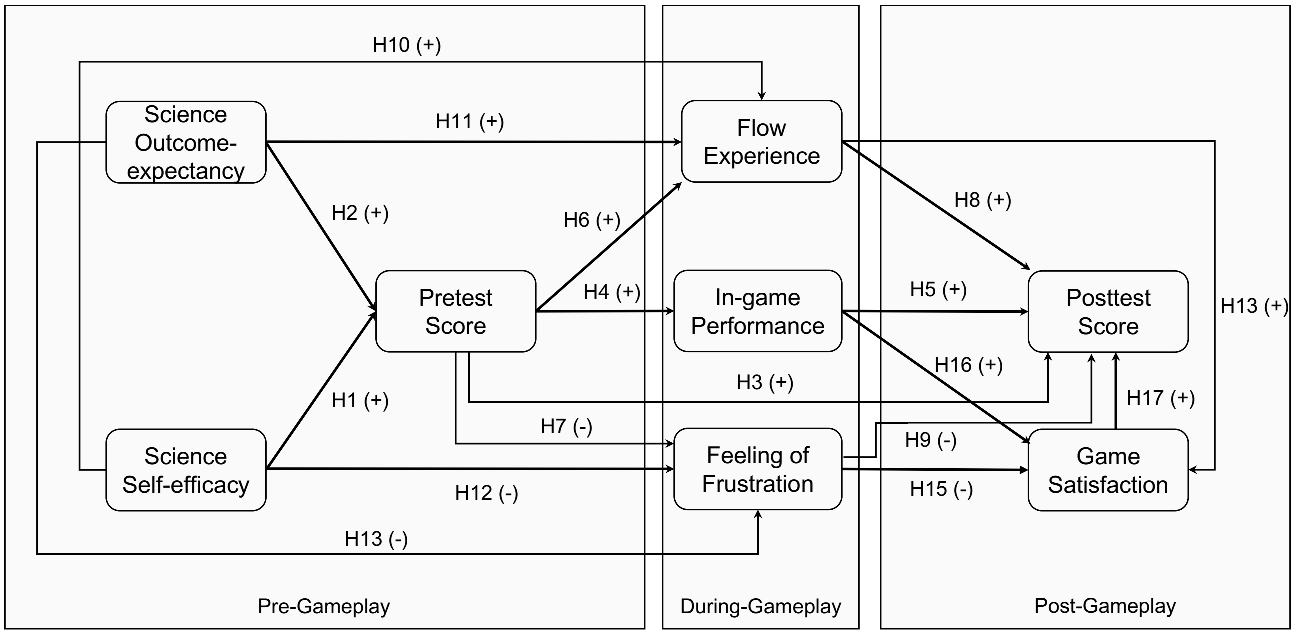

Fig. 1 Hypothesized model integrating EVT and flow theory. (+) positive impact; (−) negative impact

H_4. Genetics pretest score has a positive impact on in-game performance.

H_5. In-game performance has a positive impact on the genetics posttest score.

Flow and Prior Knowledge

Csikszentmihalyi (1990, 2014) posited that flow and frustration states depend on the relationship of one's related knowledge and skills, and the task at hand. This implies that a student's level of germane knowledge before playing the game has a relation to their being in flow or frustration states. Shute et al. (2015) demonstrated that incoming prior knowledge (pretest) has a significant positive impact on flow experience but was not significant for the feeling of frustration. Shute et al. finding on frustration conflicts with flow theory, especially in relation to the feeling of frustration that should be correlated to prior knowledge. In the present study, we investigated flow theory's prediction that higher prior knowledge should negatively influence feelings of frustration. Thus, we hypothesized the following relationships between prior knowledge and components of flow theory:

H_6. Genetics pretest score has a positive relationship with flow experience.

H_7. Genetics pretest score has a negative relationship with the feeling of frustration.

Bressler and Bodzin (2016) looked at the impact of the game-based environment on eighth-grade students' flow experience and scientific practices. They found that both students' self-reported flow experience, and their scientific practice scores improve after playing the game, suggesting a positive relationship between these two constructs. Baker et al. (2010) examined several users' cognitive-affective states while playing educational games and found that feeling frustrated produces less-than-optimal learning. Guided by these and other studies (e.g., Erhel and Jamet, 2019), we developed the following hypotheses to test to what extent flow experience and feeling of frustration during gameplay influence students' posttest scores:

H_8. Flow experience has a positive impact on the genetics posttest score.

H_9. The feeling of frustration has a negative impact on the genetics posttest score.

Self-efficacy, Outcome-Expectancy Belief, and Flow

Przybylski et al. (2010) and Sharek and Wiebe (2014) posited that one's willingness and motivation to pursue obtainable goals and overcome challenges relate to flow experience and feelings of frustration. Hung et al. (2015) examined students' science learning in tablet-PC-game–based environment and found that science self-efficacy positively correlates to flow experience and learning. This may connect to one of the ideas relating to cognitive load theory-germane cognitive load. Paas and van Merriënboer (1994) and Mayer and

Fig. 2 Screenshots from Geniventure

Moreno (2010) defined "germane cognitive load" as the cognitive effort an individual commits to learning goals. They added that the more motivated one is, the greater the cognitive processing effort involved. On the contrary, one in a frustration state would not devote effort to germane load, suggesting that motivation negatively correlates to the feeling of frustration. Based on these empirical and theoretical foundations, we proposed the following hypotheses:

H_{10}. Science self-efficacy has a positive impact on the flow experience.
H_{11}. Science outcome-expectancy belief has a positive impact on flow experience.
H_{12}. Science self-efficacy correlates negatively to the feeling of frustration.
H_{13}. Science outcome-expectancy belief correlates negatively to the feeling of frustration.

Flow, In-Game Performance and Game Satisfaction

Wiebe and colleagues (2014) showed that flow experience positively correlated to game satisfaction. They also found that feelings of frustration negatively predicted game satisfaction. Applying these findings with the prior literature on the flow and performance, Hwang et al. (2015) and Sailer et al. (2017) demonstrated that students' performance during the gameplay positively impacts their game satisfaction. In contrast, Hanus and Fox (2015) conducted a longitudinal study on the effect of gamification in the classroom. They found that students' game satisfaction did not significantly correlate with and predict their final performance. These conflicting findings may have arisen because their study was situated in a more authentic learning context than other previous studies that were mostly done over shorter time spans or lab settings. Hence, these contrasting findings warrant further examination:

H_{14}. Flow experience has a positive impact on game satisfaction.
H_{15}. The feeling of frustration has a negative impact on game satisfaction.
H_{16}. Performance in the game correlates to students' game satisfaction.
H_{17}. Game satisfaction does not have an impact on students' genetics posttest score.

Method

Research Design and Samples

The current study used a quasi-experimental, one-group pretest-posttest design and quantitative methods. We adopted this design because our prior work indicated that this digital game intervention would have a significant impact on students' learning. We then proceeded with a prediction study using path analysis where we tested our hypothesized model (Fig. 1).

A total of 394 secondary school students participated in this study. They were from seven high schools and one middle school located along the eastern seaboard of the USA. Only participants with complete in- and out-of-game data were retained for further analysis, resulting in 307 participants in the analysis. This sample size is greater than the minimum sample size of 300 suggested by Hair et al. (2019) for structural equation modeling (SEM).

The final dataset consisted of students in the seventh and ninth through twelfth grades. Most participants were in the ninth (24%) or tenth (55%) grade. Of the total sample, 47% identified as female, 46% identified as male, and the remaining 7% preferred not to indicate their gender. The participants varied in terms of ethnicity. Of those who reported their ethnicity, 51% were White, 12% were African-American, 10% were multiracial, 7% were Latinx, and 4% were Asian. The remaining 16% identified as "Other" or did not provide information about their ethnicity. Most (88%) of the participants were non-English Language Learners (ELLs) while 12% were ELLs.

The Game: Geniventure

This study involved a game-based learning environment called Geniventure in which students learned genetics concepts as they completed challenges centered on breeding dragons (aka "drakes") as part of a greater story narrative. The game was a self-paced game, and thus the length of time students took to play the game was varied (on average ranging from five to eight 90-min class periods). In the game, students completed missions consisting of individual challenges, such as manipulating alleles to match a drake's genotype to a target phenotype (Fig. 2 on the left) or choosing the correct phenotype or physical characteristic for a given allele pairing. Figure 2 shows the screenshots from the game.

Students were awarded crystals for correct submissions. The number of actions, or "moves," that students took to complete a mission was used to determine the color of the crystal that they received for that mission—blue, yellow, red, or no crystal, indicating perfect to imperfect solutions, respectively. The number of moves taken or crystals received was also used as a proxy to determine students' understanding of the genetics concepts tested in a particular mission.

The game had an adaptive support system that delivered hints to students and tracked students' performances throughout the game. The architecture of this adaptive support system was informed by the Bayes net model, used in other adaptive support systems (cf., Shute 2011). The system generated a score for each genetics concept in the game, referred to as "probability learned." When this probability learned score was below a certain threshold, students were given hints to help them complete the mission, including three levels of text-based hints along with visual cues to help them navigate to the specific area on the screen where they made an error.

Research Instrument and Data Collection

Six instruments were used in this study: science self-efficacy and outcome-expectancy belief questionnaires; flow experience and feeling of frustration questionnaires; in-game learning probability; a game satisfaction questionnaire; game experience questionnaire; and an assessment measuring students' genetics understanding.

Science Self-efficacy and Outcome Expectancy

We used the science dimension of the STEM attitudes survey developed by Unfried et al. (2015) to measure students' science self-efficacy and outcome expectancy beliefs. These questionnaires were administered before students took their pre-assessment test and consisted of four and five items, respectively, on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The items and psychometric properties are presented in Appendix 1. The Rasch analysis indicated that the instrument was valid and reliable (Boone et al. 2014; Wright and Linacre 1994). Satisfactory Cronbach's alpha values of 0.731 and 0.914 based on (DeVellis 2017) were obtained for the science self-efficacy and outcome-expectancy belief constructs, respectively.

Flow Experience and Feelings of Frustration

These two constructs were measured with two items on a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The item used to measure flow experience was "The time I spent using [The Game] just slipped away," and the item used to measure feelings of frustration was "I felt frustrated while using [The Game]." These two items were adopted from the User Engagement Scale (UES; O'Brien and Toms 2010; Wiebe et al. 2014). These two items were embedded in the exit ticket, a form of formative assessment that students took after playing [The Game] each class period. On average, students completed the game across five to eight different class periods and therefore had five to eight answers for each item. We used these answers to compute students' scores of these two variables and reliability values. The reliability value was 0.769 for flow experience and 0.609 for feelings of frustration, indicating satisfactory and acceptable values for research purposes, respectively (DeVellis 2017).

In-Game Performance/Learning Probability

An evidence-centered design (ECD, see Mislevy et al. 2012) framework for assessing student knowledge was used to analyze trace data of students' behaviors against an inventory of genetics concepts. Trace data produced by the adaptive support system were used to generate students' in-game performance scores. These scores were aggregated from scores on seven learning objectives that include 13 genetics concepts (see Appendix 2) assessed throughout the gameplay. Students' in-game performance scores ranged from 0 (no understanding) to 1 (perfect understanding) of the genetics concepts covered during the gameplay.

Game Satisfaction

Game satisfaction was also measured using the satisfaction subscale of the UES questionnaire (O'Brien and Toms 2010; Wiebe et al. 2014). The game satisfaction scale was administered post-gameplay. The scale consisted of seven five-point Likert-type items ranging from 1 (strongly disagree) to 5 (strongly agree). The items and psychometric properties are presented in Appendix 1. The Cronbach's alpha value for this scale was 0.841, indicating a reliable scale (DeVellis 2017).

Assessment Measuring Students' Genetics Understanding

Students' genetics understandings were assessed with a total of 25 multiple-choice questions developed by our research team. The same ECD process used to guide the adaptive support system development was also used to develop these items. The items then went through two rounds of revisions of earlier implementations of the game. These items tested students' understanding of genetics concepts that had been presented in the game and aligned with the learning objectives and genetics concepts listed on Appendix 2. The psychometric properties of this assessment are provided in Appendix 1. Rasch analysis indicated that the instrument was valid and reliable (see Boone et al. 2014). The Cronbach's alpha values for the pretest and posttest were 0.853 and 0.893, respectively.

Students' Affective Valence

To investigate students' affective valence in response to gameplay, we asked students to describe their experience after they played the game each class period. A single open-ended question was asked as part of the exit ticket: "Is there anything else you want to tell us about your experience with [The Game] today?" We used the data obtained from this question to answer our second research question (RQ2). Students' answers were coded based on a modified version of Russel's Core Affect Framework (Baker et al. 2010; Russell 2003). We grouped these Russel's Core Affect Framework categories into three categories: displeasure (coded with − 1), neutral (coded with 0), and pleasure (coded with 1). Examples of student quotes for each category are provided in Appendix 3. Students' responses were coded by the first author and a science education doctoral student. The inter-rater reliability was $k = 0.873$ and for each category was greater than 0.810 indicating satisfactory agreement (Cohen 1960).

Data Analysis

Rasch analysis was performed on the data for science self-efficacy, science outcome-expectancy beliefs, game satisfaction, and genetics assessment. Scores computed by Rasch analysis used the same unit of measurement, called logit, allowing the scores to be directly compared. For ease of interpretation, we also converted students' scores for some variables (science self-efficacy, science outcome-expectancy beliefs, game satisfaction, and genetics assessment) to the scale 0–100. We then ran descriptive statistics, including skewness and kurtosis, on these converted data to ensure that they were normally distributed. George and Mallery (2010) suggested that normally distributed data have skewness and kurtosis values between − 2 and 2. Next, we ran a mixed-model ANOVA to answer RQ1. We set the students' grade level in school as the covariate, given Riopel et al.'s (2020) finding that grade level moderated the impact of game-based environments on students' science achievement. We first ran the full model; when we did not find a significant interaction effect between gain score and grade, we removed the covariate from the model to obtain the main effect. Partial eta squared (η_p^2) was used to measure effect size, and 0.01, 0.06, and 0.16 designated small, medium, and large effect sizes, respectively (Cohen 1988).

Students' normalized gain scores (Hake 1998) were generated to answer RQ2. Students were divided into three groups using median-split technique based on their normalized gain scores—low, medium, and high gain scores. We then ran a chi-square test of independence to examine the association between students' game experiences, as determined by cognitive-affective states, and their learning gains. Next, we performed bivariate Pearson's and Spearman's rank correlation tests to answer RQ3. Finally, the hypothesized model in Fig. 1 was tested using SEM path analysis. The model was evaluated using the cutoff values suggested by Schreiber et al. (2006): $\chi^2/df < 3$, TLI > 0.90, CFI > 0.95, RMSEA < 0.06, and SRMR < 0.05. We then removed non-significant paths and reran the analysis, then

Table 1 Descriptive statistics

Variable	Mean (logit)	SD	Skewness	Kurtosis
Science self-efficacy	52.11	15.70	0.56	0.94
Science outcome-expectancy	52.65	21.56	0.16	0.38
Pretest score	51.73	14.87	0.99	1.16
Posttest score	64.32	16.94	0.16	− 0.52
In-game performance* (range 0–1)	0.86	0.13	− 1.96	4.87
Flow experience* (range 1–5)	3.08	0.81	− 0.03	0.01
Feeling of frustration* (range 1–5)	2.89	0.82	− 0.28	0.25
Game-satisfaction	62.17	16.19	− 0.50	4.16

*Raw mean

Table 2 Results from repeated-measures ANCOVA

Model	Variable	F	df	p	η_p^2
Model 1	Pre-post difference	5.19	[1, 298]	.023	0.017
	Grade*	0.76		.384	0.003
Model 2	Pre-post difference	165.59	[1, 306]	< .001	0.351

*Covariate, *Model 1* full model, *Model 2* without covariate

compared the hypothesized full model to the model with non-significant paths removed.

Rasch analysis was performed using WINSTEPS version 4.0.1 (Linacre 2017). SEM was performed in IBM Amos version 25.0 (Arbuckle 2017), with the remaining statistical analyses run using SPSS version 26.0 (IBM Corp. 2019).

Findings

Descriptive Statistics

Descriptive statistics for each variable are presented in Table 1. Regarding normality, for path analysis, the assumption of normality is made only for the dependent variable (Hair et al. 2019)—for this analysis, the posttest score. The non-normal independent variables are thus considered acceptable.

Pretest and Posttest Difference (RQ1)

We used mixed-model ANOVA to answer our first research question (RQ1) regarding the difference between students' conceptual understandings before and after the intervention. Two models were run: the full model with the covariate (Model 1) and a model without the covariate (Model 2). Based on the results, for Model 1, we found that grade level did not significantly impact the association between pretest and posttest score (F[1, 298] = 0.76; p = .384; η_p^2 = 0.003), indicating that students' grade level did not have a significant influence on their genetics learning in the game-based environment. We then removed grade level variable from the model to test the main effect and found a significant increase between pretest (M = 51.73; SD = 14.87) and posttest (M = 64.32; SD = 16.94) scores with a very large effect size (F[1, 306] = 165.59; p < .001; η_p^2 = 0.351). Table 2 shows all ANCOVA results.

Association Between Students' Affective Valence/Experience and Learning Gains (RQ2)

A Pearson chi-square test of independence was run to answer the second research question (RQ2) addressing the association between students' learning gains and game experiences as determined by their cognitive-affective state. In the lower scores group, 49% of students reported neutral experiences or feelings during and toward the game, 32% displeasure, and 19% pleasure. Similarly, in the medium scores group, roughly half of the students (53%) reported neutral experiences and feelings, 27% displeasure, and 19% pleasure. In the higher scores group, 56% of students voiced neutral experiences and feelings, 23% displeasure, and 21% pleasure. The chi-square test of independence found no significant association between students' learning gains and their game experiences (Pearson χ^2 = 3.22; p = .522).

Correlations Between Variables (RQ3)

Bivariate Pearson's and Spearman's rank correlation tests were run to answer RQ3, which resulted in both significant (p < .05) and non-significant correlations

Table 3 Correlation coefficients (r) computed from Pearson's and Spearman's rank correlation tests

Variable		Variable						
		(b)	(c)	(d)	(e)	(f)	(g)	(h)
Science self-efficacy	(a)	0.494**	0.274**	0.231**	0.105	0.001	− 0.108	0.158**
Science outcome-expectancy belief	(b)	–	0.304**	0.196**	0.005	0.111	− 0.090	0.193**
Pretest score	(c)		–	0.444**	0.238**	0.050	0.108	0.041
Posttest score	(d)			–	0.407**	0.140**	0.010	0.065
In-game performance[a]	(e)				–	0.068	0.017	0.038
Flow experience	(f)					–	0.004	0.245
Feeling of frustration	(g)						–	− 0.228**
Game satisfaction[a]	(h)							–

*p < 0.05; **p < 0.01; no asterisk p > 0.05

[a]Spearman's rank correlation test

Fig. 3 An example item used to measure students' understanding of genetics

3. Flying whiptails can have straight wings (W) or curly wings (w). Straight wings are dominant to curly wings. They can be dark gray (G) or light gray (g). Dark gray is dominant to light gray. You have been given some light gray flying whiptails with straight wings.

Light gray, Straight wings (ggWw)

What are the proportions of gametes (eggs or sperm) that would be made by a flying whiptail that has the genotype ggWw?

a) ½ gW and ½ gw
b) ½ gg and ½ Ww
c) ½ GW and ½ gw
d) ¼ GW, ¼ gw, ¼ gW, and ¼ Gw

($p > .05$). The most significant correlations were between science self-efficacy and outcome-expectancy belief ($r = .494$), pretest and posttest scores ($r = .444$), and in-game performance and posttest score ($r = .407$). We also found a weak but significant correlation between posttest score and flow experience ($r = .140$). Flow experience was found to be weakly correlated with science outcome-expectancy belief ($r = .111$), but this correlation was not significant ($p = .057$). Feelings of frustration were negatively and significantly correlated with game satisfaction ($r = -.228$). The complete results of the correlation tests are presented in Table 3.

Path Analysis (RQ4)

A path analysis was run to answer our fourth research question (RQ4). We tested our hypothesized model and the revised model (i.e., the model with non-significant paths removed). Removing non-significant paths improved the quality of the model according to the fit indices ($\chi^2/df = 1.877$, $p = .043$, CFI = 0.969, TLI = 0.890, and RMSEA = 0.054, for the hypothesized model, as compared with $\chi^2/df = 1.508$, $p = .093$, CFI = 0.973, TLI = 0.936, and RMSEA = 0.041 for the revised model). The improved fit indices, in particular the χ^2/df value, of the revised model

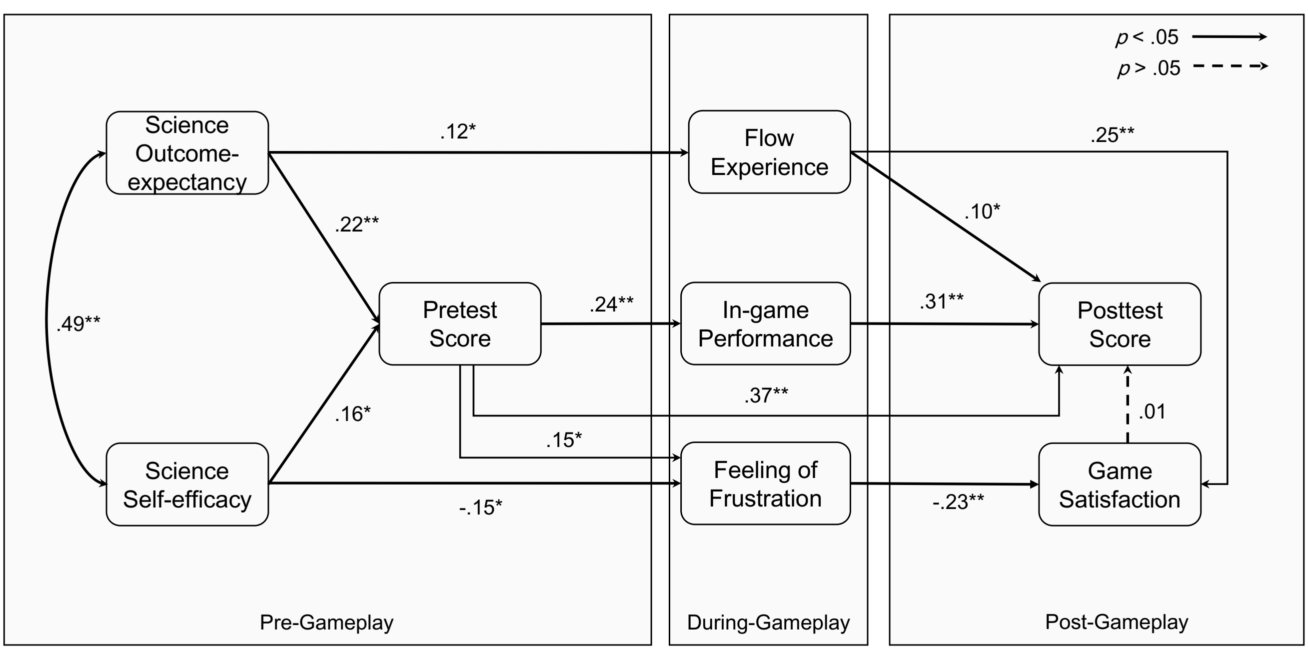

Fig. 4 Final model with standardized β values. ** $p < .01$; * $p < .05$; no asterisk $p > .05$

Table 4 Indirect effects

Indirect effect	Standardized β estimate	p value
Science expectancy beliefs → Flow experience → Posttest score	0.012	0.033
Science expectancy beliefs→ Pretest score → Posttest score	0.081	0.001
Science self-efficacy → Pretest score → Posttest score	0.060	0.011
Pretest Score → In-game performance→ Posttest score	0.075	0.001
Science self-efficacy → Feeling of frustration → Game-satisfaction	0.036	0.002
Science expectancy beliefs → Flow experience → Game-satisfaction	0.030	0.025
Pretest score → Feeling of frustration → Game-satisfaction	− 0.036	0.001
Flow experience → Game-satisfaction → Posttest score	0.003	0.735
Feeling of frustration → Game-satisfaction → Posttest score	− 0.003	0.728

indicated that the data fit the revised model better than the hypothesized model. The revised model with standardized path coefficients is visualized in Figs. 3 and 4. Each path (arrow) represents the change in Y associated with an increase in X of one standard deviation. For example, given a change of one standard deviation in science outcome-expectancy belief, pretest score improved by 0.22 standard deviations.

As shown in Fig. 4, there were two different final dependent variables—game satisfaction and posttest score—given that there was no significant path from game satisfaction to posttest score. This confirmed the findings above, which did not identify any significant association between learning gains and students' game experience. Moreover, all direct and indirect impacts from science outcome-expectancy belief and self-efficacy to posttest score were significant ($p < .05$) when they did not involve feelings of frustration. As hypothesized, pretest score, in-game performance, and flow experience were significantly predictive of posttest score. All possible indirect effects are presented in Table 4.

Discussion

In this study, we sought to examine the mechanism of students' genetics learning in a game-based environment by focusing on integrating the expectancy-value theory of achievement motivation (Wigfield and Eccles 2000) and flow theory (Csikszentmihalyi 1990). We found that the game used in this study had a significant impact on students' genetics learning as demonstrated by the significant increase in students' assessment scores, with a very large effect size. Interestingly, this result is quite different from previous empirical studies and meta-analyses of game-based learning in secondary education, which typically yielded non-significant to small effect sizes (e.g., Shute et al. 2015; Wouters et al. 2013). Critically, many previous studies did not use adaptive technologies in their games. It is possible that the real-time feedback and scaffolding via hints and the adaptive remediation provided by our game had a significant impact on learning outcomes.

It is also of note that grade level did not have a significant impact on changes in pretest and posttest scores. Several previous meta-analyses (e.g., Riopel et al. 2020; Wouters et al. 2013) found that grade level had a significant impact on students' learning in game-based environments. The alternative categorizations used in previous studies may explain this difference in findings. For instance, Riopel et al. (2020) categorized students of different grade levels into three groups representing various levels of education (primary, secondary, and college), while we used students' actual grade levels (seventh, ninth, etc.) and which constituted a relatively narrow age range. Nevertheless, our results support Riopel et al. (2020) categorization, in which none of the grade levels within the secondary level had a significant impact on students' learning and could thus be grouped into one category. Furthermore, this supports the design of the current study, especially in generating the statistical model of secondary-level students' genetics learning in game-based environments.

The results of SEM path analysis showed that 10 of 17 hypotheses were confirmed. These included the relationships between science outcome-expectancy belief, flow experience, and posttest score. As predicted, we found that science outcome-expectancy belief positively predicted flow experience during gameplay and led to an increase in posttest score. According to Csikszentmihalyi (2014), an individual needs to have a very clear goal and expectation in order to enter a flow state. Here, students' science expectations may have served that purpose, encouraging students to persist in the game and enter a flow state. One possible reason for the minimal nature of this impact is that the expectation measured in this study is not specific enough (i.e., science rather than genetics learning). Future research might examine the relationship between more specific outcome expectations (e.g., an expectation of genetics learning, rather than a broader science outcome expectation) and flow experience during gameplay.

The significant impact of flow experience on posttest score was as expected. Csikszentmihalyi (2014) indicated that when individuals are in a flow state, they usually know what they are doing as a result of receiving feedback from the activity and are therefore aware of their current state of understanding (here, genetics understanding). The regular feedback that students received during the game most likely set their baseline conceptual understanding of genetics and thus improved their performance. This may also explain why no significant path was found from pretest score to flow experience: regardless of what students knew before playing the game, the adaptive scaffolding they received during gameplay might have served as a new baseline of genetics understanding upon which they relied for improvement. Accordingly, prior knowledge may not play an important role in facilitating a flow state, at least in the context of this study.

We found other interesting results, especially the direct impact of science self-efficacy and pretest scores on students' feelings of frustration. Science self-efficacy had a significant negative relationship to feelings of frustration—that is, the lower students' science self-efficacy, the more feelings of frustration they experienced during the game. This may partially connect to a concept in flow theory regarding the alignment between individuals' perceived abilities and challenges in a game (Csikszentmihalyi 2014; Sharek and Wiebe 2015). In this study, it is possible that students were initially confident in their science abilities but realized during gameplay that their perceptions of their science skills did not align with the challenges presented in the game. If the challenges were too difficult than what students had expected, students thus experienced more frustration. In contrast, a positive impact of pretest score on feelings of frustration was identified, meaning that the higher students' pretest scores, the greater the feelings of frustration they experienced during gameplay. This result differs from both our hypothesis and Shute et al. (2015) study, which found a negative, non-significant impact of pretest score on feelings of frustration during gameplay. This finding may be explained by the expertise reversal effect (Kalyuga et al. 2003). This effect predicts that over-scaffolding for students who do not need it creates extraneous cognitive load and increased frustration (Kalyuga et al. 2003). In this case, the adaptive support system may have been delivering hints that were not needed. Thus, further work may be needed in tuning the adaptive support system's calculation of the probability guiding the hint delivery.

Finally, flow theory may offer an explanation for the non-significant impact of game satisfaction and posttest score. Csikszentmihalyi (2014) explained that the feeling of "fun" or enjoyment in the context for flow intersects with, but differs from, its counterpart in the context of satisfaction. In the context of satisfaction, enjoyment is not related to achieving goals or completing tasks in-game but is rather related to instinctual needs, like aesthetics, storyline, or other elements which also can be part of the flow experience. However, in a flow state, enjoyment involves more than instinctual aspects (Wiebe et al. 2014); the enjoyment in flow, rather, derives from "the achievement of emergent goals, that is, from one's ability to respond to opportunities in the environment that one learns about" (Csikszentmihalyi 2014, p. 159). Thus, game satisfaction may have derived from game elements that were not directly related to the posttest assessment.

Instructional Implications

Broadly, the findings of this study have demonstrated that an adaptive game-based learning environment has the potential of bridging gaps in students' prior knowledge, thus easing the burden on teachers' needs to provide individualized support. A game-based environment such as Geniventure could be used at the beginning of a genetics unit as a way of engaging students with a wide range of prior experience and getting all of the students to the same level of proficiency prior to more in-depth genetics learning activities. However, our findings on the possible negative effects of overscaffolding more advanced students points to the need to refine the learning environment to provide more challenging options.

The result of the path analysis has demonstrated the particular importance of science outcome expectancy with regard to entering a flow state during gameplay. This has practical implications, especially the degree to which teachers engage students in their outcome expectancies. Various studies across domains (e.g., health care, education, psychology) have suggested that helping individuals to explicitly identify the goals and objectives of certain activities before the activities happen evidently increases their outcome expectancies (e.g., Reesor et al. 2017; Scaduto et al. 2008; Settlege 2000). Teachers can help students identify the objectives by explicitly telling their students all the goals of the tasks or activities they are about to do. This enables students to maintain focused attention on the tasks being performed and therefore optimizing learning outcomes. Finally, the significant impact of the game-based environment on students' genetics learning suggests that the Geniventure approach to representing (visualizing) key genetics concepts was effective in supporting conceptual understanding. Thus, teachers might incorporate similar representational forms in their instruction outside of the game to help reinforce key ideas.

Conclusion, Limitations, and Directions for Future Research

In this study, we demonstrated that the outcome-expectancy value theory of achievement motivation, in connection with flow theory, provided a powerful explanatory model of genetics learning in a game-based learning environment. We found that science outcome-expectancy belief had a direct impact on students' flow experience during gameplay, whereas science self-efficacy was found to be more connected with feelings of frustration. We also found that prior knowledge did not have a significant impact on students' flow experience but was significantly associated with feelings of frustration. Finally, our results indicated that flow experience during gameplay increased students' genetics learning; however, game satisfaction did not have an impact on students' learning.

As this study had some limitations, it is important to exercise care when interpreting the results. First, we used only one item to measure flow experience and feelings of frustration. This may reduce the reliability values of the results of our study, which may in turn impact several of our findings. Future studies might address this limitation and replicate this study by adding more items—at least three for each construct (Marsh et al. 1998)—to increase its reliability. Second, the science self-efficacy and outcome-expectancy belief constructs used in this study were designed for science learning in general, not genetics learning specifically. This may partially explain why the correlations and impact were not sufficiently strong. Future research could address this issue by changing the constructs to fit more specific contexts. Moreover, future studies could add other components of expectancy-value theory, such as value constructs (Wigfield 1994), to better examine the connection between the two theories. Finally, we collected classroom observations as part of this study to confirm general fidelity of implementation (Wilson et al. 2018). However, the observation protocols were not focused on students' motivational aspects and flow experience and thus were not integrated directly into the results. Collecting data related to current studies' variables of interest from student interviews and classroom observations might help further explain and understand the mechanism of students' learning in game-based environments.

Acknowledgements We want to acknowledge all cooperating teachers who played a significant role in this study. We also thank Dolly Bounajim for helping with data cleaning. This material is based upon work supported by the United States National Science Foundation under Grant Nos. DRL-1503311. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.
Modeling Secondary Students' Genetics Learning in a Game-Based Environment: Integrating the Expectancy-Value Theory of Achievement Motivation and Flow Theory.

Ethical Approval All procedures performed in this study were in accordance with the ethical standards of the institutional and/or national research committee (North Carolina State University Protection of Human Subjects Committee, 6611) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

Appendix 1

Psychometric properties of all instruments used in this study.

Science self-efficacy

Item	Measure (Scale 100)	Infit MNSQ	Outfit MNSQ	Cronbach's alpha if item deleted	Person reliability	Item reliability
I am sure of myself when I do science	47.07	0.88	0.91	0.676	0.72	0.96
I know I can do well in science	39.35	0.74	0.68	0.588		
I can handle most subjects well, but I cannot do a good job with science	44.55	1.39	1.36	0.737		
I am sure I could do advanced work in science	50.87	1.02	0.99	0.675		

Science outcome expectancy

Item	Measure (Scale 100)	Infit MNSQ	Outfit MNSQ	Cronbach's alpha if item deleted	Person reliability	Item reliability
I would consider a career in science	51.78	1.23	1.23	0.903	0.88	0.96
I expect to use science when I get out of school	41.56	1.08	1.02	0.897		
Knowing science will help me earn a living	42.55	0.89	0.90	0.891		
I will need science for my future work	46.49	0.83	0.83	0.889		
Science will be important to me in my life's work	47.69	0.87	0.86	0.894		

Genetics assessment

Item	Measure (Scale 100)	Infit MNSQ	Outfit MNSQ	Cronbach's alpha if item deleted	Person Reliability	Item reliability
Item1	45.97	0.84	0.68	0.866	0.83	0.98
Item2	45.64	0.90	0.80	0.867		
Item3	38.15	0.82	0.58	0.867		
Item4	34.72	0.82	0.56	0.868		
Item5	49.15	1.03	1.04	0.869		
Item6	45.43	1.08	1.30	0.870		
Item7	43.55	1.19	1.30	0.873		
Item8	49.09	0.75	0.63	0.863		
Item9	45.01	0.89	0.78	0.867		
Item10	56.99	0.94	0.95	0.867		
Item11	55.34	0.86	0.83	0.866		
Item12	55.12	0.71	0.63	0.862		
Item13	40.72	0.94	0.81	0.869		
Item14	62.32	1.05	1.15	0.868		
Item15	45.34	0.99	1.05	0.869		
Item16	53.54	0.98	0.97	0.867		
Item17	37.99	0.93	0.65	0.869		
Item18	43.66	1.04	1.05	0.869		
Item20	50.67	1.00	0.98	0.867		
Item22	56.85	1.26	1.45	0.872		
Item23	45.54	1.04	1.00	0.869		
Item24	59.58	1.32	1.41	0.873		
Item25	72.66	1.07	1.40	0.871		
Item26	58.30	1.36	1.46	0.873		
Item28	46.30	1.09	1.35	0.870		

Game satisfaction

Item	Measure (Scale 100)	Infit MNSQ	Outfit MNSQ	Cronbach's alpha if item deleted	Person reliability	Item reliability
Using Geniventure was worthwhile	49.22	0.94	1.05	0.817	0.71	0.55
I consider my experience a success	46.20	1.10	0.92	0.833		
My experience was rewarding	49.08	1.01	0.95	0.810		
I would recommend Geniventure to my classmates	50.25	0.91	0.81	0.806		
Geniventure made me more curious about genetics	49.72	1.12	1.18	0.826		
I felt involved in this experience	48.70	0.75	0.70	0.816		
This experience was fun	48.04	1.11	1.07	0.822		

Appendix 2

Lists of Learning Objectives and Genetics Concepts Associated with Geniventure and Genetics Assessment.

Learning objectives

LG 1. There are predictable correlations between an organism's genes and its traits.

LG 2. Genetic information is passed to an individual from both its parents via their gametes.

LG 3. Processes of inheritance involve randomized events that produce predictable patterns in offspring populations.

LG 4. Genes are instructions for constructing proteins.

LG 5. Proteins carry out a variety of functions in cells.

LG 6. Protein function is a result of protein structure.

Genetics concepts

1. Sex determination (LG1.A3)
2. Simple dominance (LG1.C2a)
3. Recessive traits (LG1.C2b)
4. X linked genes (LG1.C2c)
5. Polyallelic (LG1.C2d)
6. Incomplete dominance (LG1.C3)
7. Genotype-to-phenotype mapping (LG1.P1)
8. Phenotype-to-genotype mapping (LG1.P2)
9. Epistasis (LG1.C4a)
10. Gamete selection (LG2.P1)
11. Parent genotypes (LG3.P1)
12. Patterns in offspring (LG3.P3)
13. Test cross (LG3.P4)

Appendix 3

The coding scheme for students' game experiences (Adapted from Baker et al. 2010)

Code for category	Category	Subcategory	Definition	Example	Dimension
1	Displeasure/negative feeling/experience ($k = 0.857$)	Boredom	When participants expressed any weary feelings and no interest	"boring" [Student ID 234423]	High frustration due to interface and control problems
		Frustration	When participants expressed any dissatisfaction, annoyance, and other frustrating feelings/experiences	"The activities when you have to click the triangles to change the dragon color is very unclear and annoying" [Student ID 251491]	
		Confusion	When participants expressed any noticeable lack of understanding	"On level 4.1 on the second gem, the amount of armor on the target dragon needs to be more apparent because I kept getting it wrong because I couldn't tell how much armor was on the target dragon." [Student ID 252074] "Level 3.2 was not easy to understand. I really struggled with figuring out what I was supposed to do." [Student ID 246228]	
		Difficult	When participants expressed that the mission or challenge was too hard	"It was quite difficult to get a blue crystal at this point." [Student ID 241121] "it took me a full hour to do the very last task. really hard to understand and no hints. would appreciate a help button for when you are stuck" [Student ID 235709]	
		Display or other system problems	When participants expressed that they dissatisfied with display or interface, unclear directions and when they experienced any error or game issues	"The graphics were a little slow and some things were hard using the trackpad on the chromebooks, but we didn't have another option" [Student ID 172936]	
2	Neutral ($k = 0.841$)	Neutral	No apparent feeling or emotion, including "No," "Nope," and "Easy"	"There is nothing else I want to say about my experience with [game]." [Student ID 247450]	Most students said "No"
		Surprise	When participants expressed amazement or wonder from the unexpected	"the thing is how the dragon has horn" [Student ID 235814]	
3	Pleasure/positive feeling/experience ($k = 0.922$)	Delight	When participants expressed any satisfaction, including with pleasure on visuals, difficulty, and other experiences. Also, this may include "Yes" and "OK"	"Really fun to learn and play." [Student ID 234176] "…it was fun and challenging." [Student ID 216324] "my experience was good because I learn about genes." [Student ID 235803]	Students enjoyed playing the game and learned something from the game
		Engaged concentration	When participants expressed interest in the game resulting from the involvement in the game activity	"It was fun. Also it was a little challenging at first then I understood it." [Student ID 234425]	
NR			When participants' statements are not interpretable and do not fall under the above criteria	"I'm a reptiliologist" [Student ID 238430] "I don´t like the survey." [Student ID 235590]	

References

Abraham, J., & Barker, K. (2015). An expectancy-value model for sustained enrolment intentions of senior secondary physics students. *Research in Science Education, 45*(4), 509–526. https://doi.org/10.1007/s11165-014-9434-x.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers and Education, 53*(1), 74–85. https://doi.org/10.1016/j.compedu.2008.12.020.

Arbuckle, J. L. (2017). *Amos (Version 25.0)* [Computer Program]. Chicago, IL: SPSS.

Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003.

Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: W. H. Freeman.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 15*(1), 1–40. https://doi.org/10.1023/A:1021302408382

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht, Netherlands: Springer.

Bressler, D. M., & Bodzin, A. M. (2016). Investigating flow experience and scientific practices during a mobile serious educational game. *Journal of Science Education and Technology, 25*(5), 795–805. https://doi.org/10.1007/s10956-016-9639-z.

Burak, S. (2014). Motivation for instrument education: a study with the perspective of expectancy-value and flow theories. *Eurasian Journal of Educational Research*, *55*, 123–136. https://doi.org/10.14689/ejer.2014.55.8.

Cheng, M. T., & Annetta, L. (2012). Students' learning outcomes and learning experiences through playing a Serious Educational Game. *Journal of Biological Education, 46*(4), 203–213. https://doi.org/10.1080/00219266.2012.688848.

Cheng, M. T., She, H. C., & Annetta, L. A. (2015). Game immersion experience: its hierarchical structure and impact on game-based science learning. *Journal of Computer Assisted Learning, 31*(3), 232–253. https://doi.org/10.1111/jcal.12066.

Cheng, M. T., Su, T., Huang, W. Y., & Chen, J. H. (2014). An educational game for learning human immunology: What do students learn and how do they perceive?. *British Journal of Educational Technology*, *45*(5), 820-833. https://doi.org/10.1111/bjet.12098

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Csikszentmihalyi, M. (1988). *The flow experience and its significance for human psychology.*In M. Csikszentmihalyi & I. S. Csikszentmihalyi (Eds.),*Optimal experience: Psychological studies of flow in consciousness* (p. 15–35). Cambridge University Press

Csikszentmihalyi, M. (1990). *Flow: the psychological of optimal experience*. New York, NY: Harper & Row.

Csikszentmihalyi, M. (2014). *Applications of flow in human development and education: the collected works of Mihaly Csikszentmihalyi*. Springer.

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., et al. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education, 28*(2), 152–193. https://doi.org/10.1007/s40593-017-0152-1.

DeVellis, R. F. (2017). *Scale development: theory and applications* (4th ed.). Los Angeles, CA: Sage.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology, 53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153.

Erhel, S., & Jamet, E. (2019). Improving instructions in educational computer games: exploring the relations between goal specificity, flow experience and learning outcomes. *Computers in Human Behavior*, *91*(May 2018), 106–114.https://doi.org/10.1016/j.chb.2018.09.020.

Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. (2014). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Educational Technology & Society*, *17*(1), 42–53. https://www.jstor.org/stable/jeductechsoci.17.1.42.

George, D. & Mallery, P. (2010). *SPSS for Windows step by step: A simple guide and reference 17.0 Update. 10th Edition*. Boston, MA: Pearson.

Gericke, N., & Wahlberg, S. (2013). Clusters of concepts in molecular genetics: a study of Swedish upper secondary science students understanding. *Journal of Biological Education, 47*(2), 73–83. https://doi.org/10.1080/00219266.2012.716785.

Guo, J., Marsh, H. W., Parker, P. D., Morin, A. J., & Dicke, T. (2017). Extending expectancy-value theory predictions of achievement and aspirations in science: dimensional comparison processes and expectancy-by-value interactions. *Learning and Instruction, 49,* 81–91. https://doi.org/10.1016/j.learninstruc.2016.12.007.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate data analysis* (Eight). Hampshire, UK: Cengage Learning.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: a six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*(1), 64–74. https://doi.org/10.1119/1.18809.

Hanus, M. D., & Fox, J. (2015). Assessing the effects of gamification in the classroom: a longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance. *Computers & Education, 80,* 152–161. https://doi.org/10.1016/j.compedu.2014.08.019.

Hone, K. (2006). Empathic agents to reduce user frustration: the effects of varying agent characteristics. *Interacting with Computers, 18*(2), 227–245. https://doi.org/10.1016/j.intcom.2005.05.003.

Hung, C. Y., Sun, J. C. Y., & Yu, P. T. (2015). The benefits of a challenge: student motivation and flow experience in tablet-PC-game-based learning. *Interactive Learning Environments, 23*(2), 172–190. https://doi.org/10.1080/10494820.2014.997248.

Hwang, G. J., Chiu, L. Y., & Chen, C. H. (2015). A contextual game-based learning approach to improving students' inquiry-based learning performance in social studies courses. *Computers & Education, 81,* 13–25. https://doi.org/10.1016/j.compedu.2014.09.006.

IBM Corp. (2019). *IBM SPSS Statistics for Windows, Version 26.0* [Computer Program]. Armonk, NY: IBM Corp.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31. https://doi.org/10.1207/S15326985EP3801_4.

Kim, B., Pathak, S. A., Jacobson, M. J., Zhang, B., & Gobert, J. D. (2015). Cycles of exploration, reflection, and consolidation in model-based learning of genetics. *Journal of Science Education and Technology, 24*(6), 789–802. https://doi.org/10.1007/s10956-015-9564-6.

Lauermann, F., Tsai, Y. M., & Eccles, J. S. (2017). Math-related career aspirations and choices within Eccles et al.'s expectancy–value theory of achievement-related behaviors. *Developmental Psychology, 53*(8), 1540–1559. https://doi.org/10.1037/dev0000367.

Li, M. C., & Tsai, C. C. (2013). Game-based learning in science education: a review of relevant research. *Journal of Science Education and Technology, 22*(6), 877–898. https://doi.org/10.1007/s10956-013-9436-x.

Linacre, J. M. (2017). *Winsteps (Version 4.0.1)* [Computer Program]. Available from http://www.winsteps.com/index.html.

Marbach-Ad, G., Rotbain, Y., & Stavy, R. (2008). Using computer animation and illustration activities to improve high school students' achievement in molecular genetics. *Journal of Research in Science Teaching, 45*(3), 273–292. https://doi.org/10.1002/tea.20222.

Marsh, H. W., Hau, K. T., Balla, J. R., & Grayson, D. (1998). Is more ever too much? The number of indicators per factor in confirmatory factor analysis. *Multivariate Behavioral Research, 33*(2), 181–220. https://doi.org/10.1207/s15327906mbr3302_1.

Mayer, R., & Moreno, R. (Ed.). (2010). Cognitive load theory: historical development and relation to other theories. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (p. 9–28). Cambridge University Press.https://doi.org/10.1017/CBO9780511844744.003.

Melhárt, D. (2018). Towards a comprehensive model of mediating frustration in videogames. *Game Studies*, *18*(1). Retrieved from http://gamestudies.org/1801/articles/david_melhart.

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., Frezzo, D. C., & West, P. (2012). Three things game designers need to know about assessment. In D. Ifenthaler, D. Eseryel, & X. Ge (Eds.), *Assessment in Game-Based Learning* (pp. 59–81). New York, NY: Springer.

Nakamura, J., & Csikszentmihalyi, M. (2014). The concept of flow. In M. Csikszentmihalyi (Ed) *Flow and the foundations of positive psychology* (pp. 239–263). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9088-8_16.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology, 59*(6), 938–955. https://doi.org/10.1002/asi.20801.

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50–69. https://doi.org/10.1002/asi.21229.

Paas, F. G., & van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: a cognitive-load approach. *Journal of educational psychology, 86*(1), 122–133. https://doi.org/10.1037/0022-0663.86.1.122.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*(4), 543–578. https://doi.org/10.3102/00346543066004543.

Pajares, F. (1997). Current directions in self-efficacy research. *Advances in Motivation and Achievement, 10*(149), 1–49.

Przybylski, A. K., Rigby, C. S., & Ryan, R. M. (2010). A motivational model of video game engagement. *Review of General Psychology, 14*(2), 154–166. https://doi.org/10.1037/a0019440.

Reesor, L., Vaughan, E. M., Hernandez, D. C., & Johnston, C. A. (2017). Addressing outcomes expectancies in behavior change. *American Journal of Lifestyle Medicine, 11*(6), 430–432. https://doi.org/10.1177/1559827617722504.

Riopel, M., Nenciovici, L., Potvin, P., Chastenay, P., Charland, P., Sarrasin, J. B., & Masson, S. (2020). Impact of serious games on science learning achievement compared with more conventional instruction: an overview and a meta-analysis. *Studies in Science Education*, 1-46.https://doi.org/10.1080/03057267.2019.1722420.

Rotbain, Y., Marbach-Ad, G., & Stavy, R. (2008). Using a computer animation to teach high school molecular biology. *Journal of Science Education and Technology, 17*(1), 49–58. https://doi.org/10.1007/s10956-007-9080-4.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172. https://doi.org/10.1037/0033-295X.110.1.145.

Sailer, M., Hense, J. U., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: an experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69,* 371–380. https://doi.org/10.1016/j.chb.2016.12.033.

Scaduto, A., Lindsay, D., & Chiaburu, D. S. (2008). Leader influences on training effectiveness: motivation and outcome expectation processes. *International Journal of Training and Development, 12*(3), 158–170. https://doi.org/10.1111/j.1468-2419.2008.00303.x.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: a review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338.

Settlage, J. (2000). Understanding the learning cycle: Influences on abilities to embrace the approach by preservice elementary school teachers. *Science Education, 84*(1), 43–50.

Sharek, D., & Wiebe, E. (2014). Measuring video game engagement through the cognitive and affective dimensions. *Simulation & Gaming, 45*(4–5), 569–592. https://doi.org/10.1177/1046878114554176.

Sharek, D., & Wiebe, E. (2015). Investigating real-time predictors of engagement: implications for adaptive videogames and online training. *International Journal of Gaming and Computer-Mediated Simulations, 7*(1), 21–36. https://doi.org/10.4018/IJGCMS.2015010102.

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, NC: Information Age Publishers.

Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., et al. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. *Computers and Education, 86,* 224–235. https://doi.org/10.1016/j.compedu.2015.08.001.

Su, C. H. (2016). The effects of students' motivation, cognitive load and learning anxiety in gamification software engineering education: a structural equation modeling study. *Multimedia Tools and Applications, 75*(16), 10013–10036. https://doi.org/10.1007/s11042-015-2799-7.

Unfried, A., Faber, M., Stanhope, D. S., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, engineering, and math (S-STEM). *Journal of Psychoeducational Assessment, 33*(7), 622–639. https://doi.org/10.1177/0734282915571160.

Uçar, F. M., & Sungur, S. (2017). The role of perceived classroom goal structures, self-efficacy, and engagement in student science achievement. *Research in Science & Technological Education, 35*(2), 149–168. https://doi.org/10.1080/02635143.2017.1278684.

Vogel, J. J., Greenwood-Ericksen, A., Cannon-Bowers, J., & Bowers, C. A. (2006). Using virtual reality with and without gaming attributes for academic achievement. *Journal of Research on Technology in Education, 39*(1), 105–118. https://doi.org/10.1080/15391523.2006.10782475.

Wiebe, E., Unfried, A., & Faber, M. (2018). The relationship of STEM attitudes and career interest. *EURASIA Journal of Mathematics, Science and Technology Education, 14*(10), 1-17. https://doi.org/10.29333/ejmste/92286.

Wiebe, E. N., Lamb, A., Hardy, M., & Sharek, D. (2014). Measuring engagement in video game-based environments: Investigation of the User Engagement Scale. *Computers in Human Behavior, 32,* 123–132. https://doi.org/10.1016/j.chb.2013.12.001.

Wigfield, A. (1994). Expectancy-value theory of achievement motivation: A developmental perspective. *Educational Psychology Review, 6*(1), 49–78. https://doi.org/10.1007/BF02209024.

Wigfield, A., & Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary Educational Psychology, 25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015.

Wilson, C., Reichsman, F., Mutch-Jones, K., Gardner, A., Marchi, L., Kowalski, S., & Lord, T. (2018). Teacher implementation and the impact of game-based science curriculum materials. *Journal of Science Education and Technology, 27*(4), 285–305. https://doi.org/10.1007/s10956-017-9724-y.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105*(2), 249–265. https://doi.org/10.1037/a0031311.

Wright, B. D., & Linacre, J. M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions, 8*(3), 370.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.