Carnegie Mellon University

Human-Computer Interaction Institute

Moving Beyond Test Scores: Investigating the Effectiveness of a Digital Learning Game through Learning Analytics

Huy A Nguyen, Xinying Hou, John Stamper, Bruce M McLaren

Presented by:

Huy A. Nguyen

Xinying Hou

Digital learning games (DLGs) are...

> Powerful tool to promote learning by engaging students with an interactive game environment

Initial concerns about DLGs' effectiveness

(Honey & Hilton, 2011; Mayer, 2014)

> Recently, More research addresses this issue by showing students' **learning gains from pretest to posttest** in rigorous randomized experiments.
>
> **(e.g., Chen et al., 2015; Erhel & Janet, 2016; McLaren et al., 2017; Sawyer et al., 2017; Ninaus et al., 2017)**

Initial concerns about DLGs' effectiveness

(Honey & Hilton, 2011; Mayer, 2014)

Recently, More research addresses this issue by showing students' **learning gains from pretest to posttest** in rigorous randomized experiments.

(e.g., Chen et al., 2015; Erhel & Janet, 2016; McLaren et al., 2017; Sawyer et al., 2017; Ninaus et al., 2017)

Meta-analysis of 69 learning games:

Game conditions > non-game conditions

Augmented game designs > Standard game designs

in promoting learning

(Clark et al., 2016)

While this prior research has demonstrated that digital learning games can enhance learning...

The next step is to examine how they do so

> In particular, even though the common measures of **pretest and posttest scores** are necessary to evaluate students' transferable learning

The next step is to examine how they do so

In particular, even though the common measures of **pretest and posttest scores** are necessary to evaluate students' transferable learning

> They are **inadequate** to address many questions about **how learning takes place during the game**

The next step is to examine how they do so

In particular, even though the common measures of **pretest and posttest scores** are necessary to evaluate students' transferable learning

They are **inadequate** to address many questions about **how learning takes place during the game**

For example:
Did students **get just enough practice** from the game?
Did students get **more practice than necessary**?
How does **in-game learning correlate** with **test performance**?

Research Questions

RQ1: How well did students learn in the game?

RQ2: Which factors affected students' post-game performance?

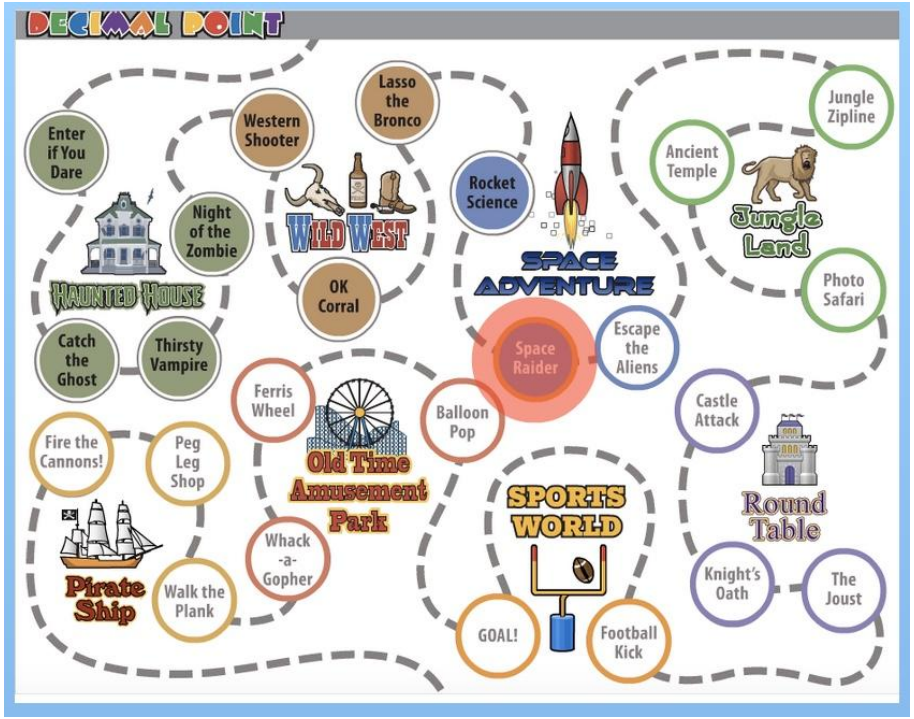
RQ3: Which factors affected students' enjoyment ratings after game play?

Decimal Point: A math digital learning game for middle-school students

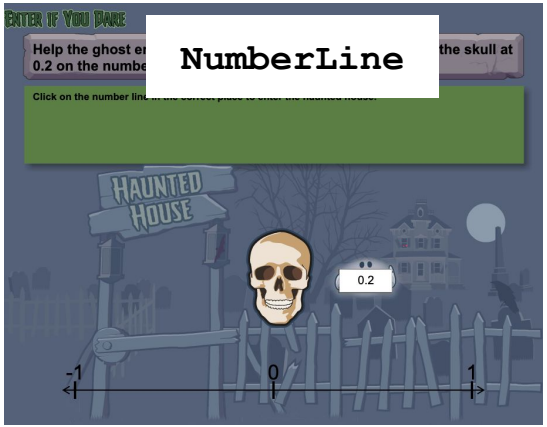
Decimal numbers and misconceptions

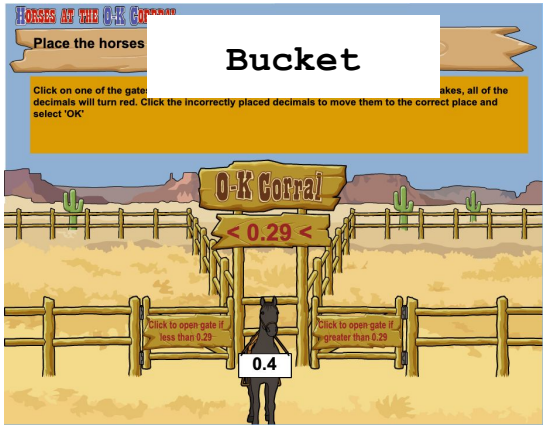
Amusement park metaphor

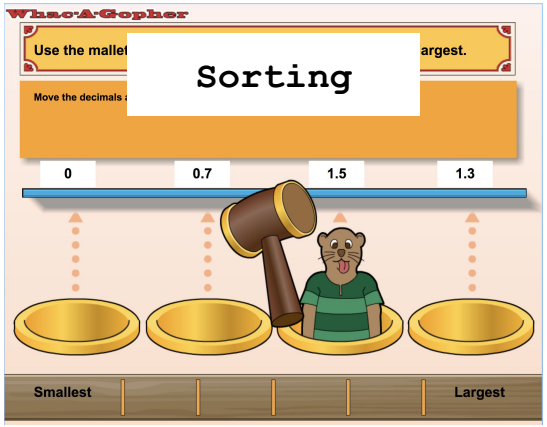
6 theme areas
24 mini-games

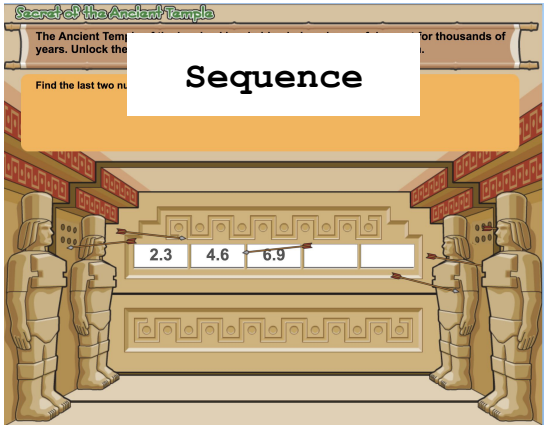


The Five Mini-game Types in Decimal Point









Procedure: Classroom Experiment

159 fifth and sixth grade students from **3** middle schools

(Hou et al., 2020)

Pre-intervention	Pretest	Demographic	Game survey
Intervention (up to 3 days in math class time)		Game play	
Immediate after intervention	Evaluation survey		Posttest
One week after intervention		Delayed posttest	

Key Measurements

Learning: In-game Learning:

Post-game learning performance:

Key Measurements

Learning: In-game Learning:

Final Mastery Probabilities

Key Measurements

Learning: In-game Learning:

Final Mastery Probabilities

- The KCs are the five mini-game types
- Tracked students' learning progress in these KCs by Bayesian Knowledge Tracing (BKT) **(Yudelson et al, 2013)**.
- Mastery threshold: 0.9.

Key Measurements

Learning:

In-game Learning:

Final Mastery Probabilities

- The KCs are the five mini-game types
- Tracked students' learning progress in these KCs by Bayesian Knowledge Tracing (BKT) **(Yudelson et al, 2013)**.
- Mastery threshold: 0.9.

Post-game learning performance:

Posttest scores and Delayed posttest scores

- Each test consisted of 43 items, for a total of 52 points.
- e.g., "is a longer decimal larger than a shorter decimal?"

Key Measurements

Learning:

In-game Learning:

Final Mastery Probabilities

Test performance after the game:

Posttest scores and Delayed posttest scores

Enjoyment:

Self-reported Enjoyment in evaluation surveys

- Per-student average Likert scores
- Achievement emotion, Game engagement, Affective engagement

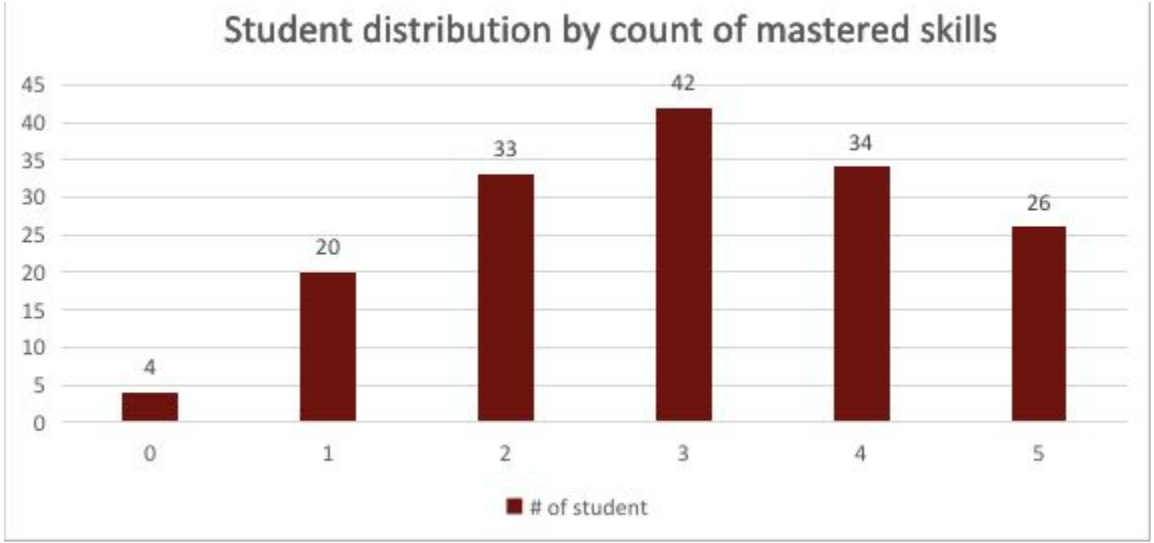
Key Measurements

Demographic Information:

Pre-intervention surveys	
Dimension (item count)	**Example statement**
Decimal efficacy (3) [44]	I can do an excellent job on decimal number math assignments.
Computer efficacy (3) [31]	I know how to find information on a computer.
Identification agency (2) [50]	I work on my classwork because I want to learn new things.
Intrinsic agency (2) [50]	I work on my classwork because I enjoy doing it.
External agency (3) [50]	I work on my classwork so the teacher won't be upset with me.
Perseverance (3) [12]	Setbacks don't discourage me. I don't give up easily.
Math utility (3) [13]	Math is useful in everyday life.
Math interest (2) [14]	I find working on math to be very interesting.
Expectancy (1) [23]	I plan to take the highest level of math available in high school.

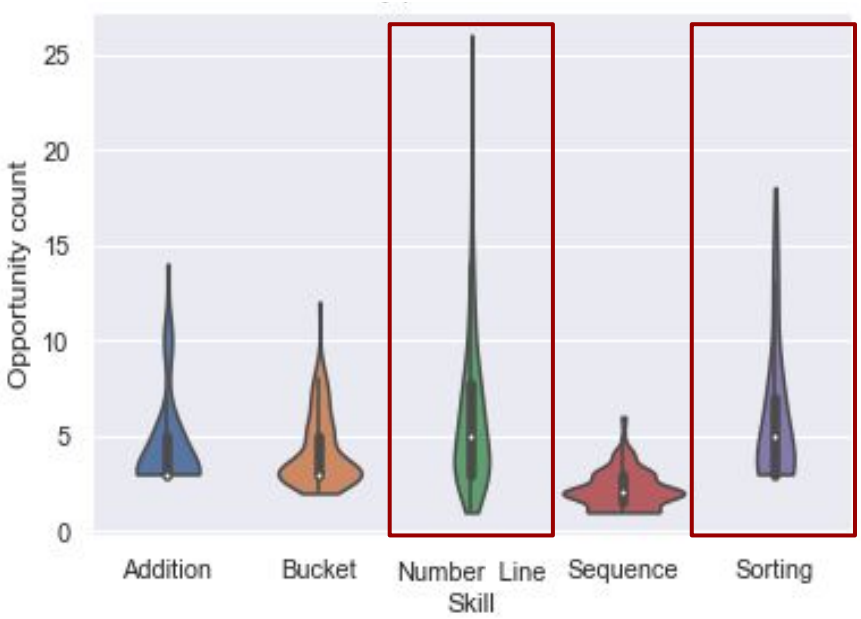
RQ1: How well did students learn in the game?

How did students master the five skills in the game?



RQ1: How well did students learn in the game?

How many opportunities did each student who mastered a skill take to reach mastery in that skill?



- **An opportunity**: one complete decimal question in a mini-game type
- `Number Line` and `Sorting` took the longest to master, at around 5 opportunities on average.
- One student even needed 26 opportunities to master `Number Line`.

RQ1: How well did students learn in the game?

How did students regulate their learning?

After mastering a skill, did they tend to

- continue practicing the same skill?
- switch to a different skill?

RQ1: How well did students learn in the game?

How well did students regulate their learning?

After mastering a skill, did they tend to

- **continue practicing the same skill?**
- switch to a different skill?

→ between **20-80%** of a student's practice opportunities in a skill could be considered as **over-practice**

RQ2: Which factors affected students' post-game performance?

RQ2: Which factors affected students' post-game performance?

19 starting features

- pretest score
- decimal efficacy
- gender
- computer efficacy
- identification agency
- intrinsic agency
- external agency
- perseverance
- utility
- math interest
- expectancy
- final in-game mastery probabilities of the five skills (`Addition, Bucket, Sequence, Number Line, Sorting`)
- total opportunity count
- over-practice opportunity count
- total incorrect answer counts.

RQ2: Which factors affected students' post-game performance?

19 starting features

- **Forward feature selection** with linear regression
  → feature subset X_1 , X_2 , ..., X_k with the lowest cross-validated **mean-squared error**

RQ2: Which factors affected students' post-game performance?

19 starting features

- **Forward feature selection** with linear regression
  → feature subset $X_1, X_2, ..., X_k$ with the lowest cross-validated **mean-squared error**
- Linear regression on the whole dataset

  Posttest / Delayed Posttest Score ~ $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_k X_k$

- Evaluate adjusted **R^2**

RQ2: Which factors affected students' post-game performance?

	Posttest	Delayed Posttest
Selected features	Pretest score, `Bucket` mastery, `Sorting` mastery	Pretest score, `Bucket` mastery, `Sorting` mastery, `Number Line` mastery, gender
Overall performance	MSE = 26.167, **adjusted R^2 = 0.735**	MSE = 24.218, **adjusted R^2 = 0.747**

RQ3: Which factors affected students' enjoyment ratings after game play?

19 starting features

- **Forward feature selection** with linear regression
  → feature subset X_1 , X_2 , ..., X_k with the lowest cross-validated **mean-squared error**
- Linear regression on the whole dataset

  Enjoyment Rating ~ $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$

- Evaluate adjusted **R^2**

RQ3: Which factors affected students' enjoyment ratings after game play?

	Achievement Emotion	**Game Engagement**	**Affective Engagement**
Selected features	computer efficacy, identification agency, intrinsic agency, math interest, pretest score, total opportunity count	math interest, computer efficacy, gender	decimal efficacy, gender, intrinsic agency, `Sorting` mastery, `Bucket` mastery, total incorrect attempt count, identification agency
Overall performance	MSE = 0.520, **adjusted R^2 = 0.386**	MSE = 0.602, **adjusted R^2 = 0.225**	MSE = 0.660, **adjusted R^2 = 0.218**

Discussion 1: Investigating in-game learning

> `Sorting` and `Number Line` are the most difficult skills

Discussion 1: Investigating in-game learning

`Sorting` and `Number Line` are the most difficult skills

> **Possible Reason:**
> Students may shift from one misconception to another
> **(Van Dooren et al., 2004)**

Discussion 1: Investigating in-game learning

`Sorting` and `Number Line` are the most difficult skills

Possible Reason:
Students may shift from one misconception to another
(Van Dooren et al., 2004)

These results call for:

- more refined knowledge tracing to track students' misconception
- more in-game instructional support: explanatory feedback, error messages

Discussion 1: Investigating in-game learning

> Only 26/159 students mastered all 5 skills. 20-80% of practice opportunities were over-practice.

Discussion 1: Investigating in-game learning

Only 26/159 students mastered all 5 skills. 20-80% of practice opportunities were over-practice.

Possible Reason:
Game environment did not indicate when mastery is reached

Discussion 1: Investigating in-game learning

Only 26/159 students mastered all 5 skills. 20-80% of practice opportunities were over-practice.

Possible Reason:
Game environment did not indicate when mastery is reached

This result suggests:
Provide skill mastery indicator, and suggest which skill to practice next **(Long & Aleven, 2017)**

Discussion 2: Investigating factors related to test performance

> Pretest score, `Sorting` mastery and `Bucket` mastery are **significant predictors** of both **posttest and delayed posttest scores**.
> `Number Line` is a **significant predictor** of **delayed posttest scores**.

Discussion 2: Investigating factors related to test performance

Pretest score, `Sorting` mastery and `Bucket` mastery are **significant predictors** of both **posttest and delayed posttest scores**.
`Number Line` is a **significant predictor** of **delayed posttest scores**.

These results suggest:

- These KCs are the differentiator KCs in the test
- Future work: more instructional support for decimal comparison and `Number Line` problems.

Discussion 3: Investigating factors related to enjoyment

Enjoyment models did not perform as well as learning models

- $R^2 \cong 0.2$ for predicting game engagement and affective engagement, $\cong 0.4$ for achievement emotion

Discussion 3: Investigating factors related to enjoyment

Enjoyment models did not perform as well as learning models

- $R^2 \cong 0.2$ for predicting game engagement and affective engagement, $\cong 0.4$ for achievement emotion

Possible Reason:
Lack of appropriate features in our data
E.g., interaction traces, decision making time, social engagement profile
(Bouvier et al, 2014; Riemer & Schrader, 2016; Ruiperez-Valiente et al, 2020)

Conclusion

- `Sorting` and `Number Line` are important skills for test performance, but students **required more instructional support** to effectively master them.
- Very few students mastered all five decimal skills in the game, while **the majority engaged in over-practice**.
- There is a **need for more fine-grained features** to capture students' enjoyment in the game.

Broad Vision

Digital learning games can optimize student learning while retaining their core value as a playful environment, where players are free to exercise their agency.

Thank you!

Carnegie Mellon University

Human-Computer Interaction Institute

Huy A. Nguyen

Xinying Hou

John Stamper

Bruce M. McLaren

For more information: http://tiny.cc/DecimalPoint

Acknowledgements. This work was supported by NSF Award #DRL-1238619. The opinions expressed are those of the authors and do not represent the views of NSF. Thanks to Jodi Forlizzi, Michael Mogessie Ashenafi, Scott Herbst, Craig Ganoe, Darlan Santana Farias, Rick Henkel, Patrick B. McLaren, Grace Kihumba, Kim Lister, Kevin Dhou, John Choi, and Jimit Bhalani, all of whom made important contributions to the development of and early experimentation with the Decimal Point game.

Discussion 2: Investigating factors related to test performance

> `Number Line` has significant association with posttest but not delayed posttest score

Discussion 2: Investigating factors related to test performance

`Number Line` is significant association with posttest but not delayed posttest score

`Number Line` also took the most opportunities to master

Discussion 2: Investigating factors related to test performance

`Number Line` is significant association with posttest but not delayed posttest score

`Number Line` also took the most opportunities to master

> `Number Line` problems could be at a desirable difficulty level
> **(Yue et al, 2013)**

Discussion 2: Investigating factors related to test performance

`Number Line` is significant association with posttest but not delayed posttest score

`Number Line` also took the most opportunities to master

`Number Line` problems could be at a desirable difficulty level
(Yue et al, 2013)

These results suggest:

- `Number Line` and `Sorting` are the differentiator KCs in the test
- Future work: more in-game support for these two skills

Conclusion

- Identified a trend of females outperforming males in the delayed posttest, which should be investigated on a larger sample size.
- Students' achievement emotion can be reasonably captured by their level of computer efficacy, learning motivation, prior knowledge and number of mini-game rounds.

RQ1: How well did students learn in the game?

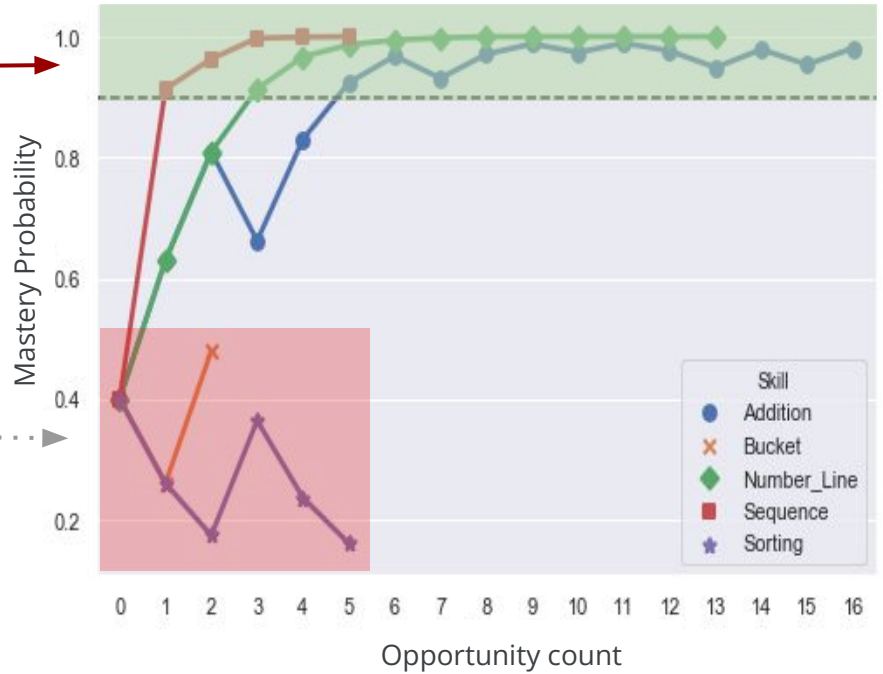
How well did students regulate their learning?

Continue practicing mastered skills:

`Sequence, Number Line, Addition`

Ignore the other two unmastered skills

`Bucket, Sorting`



Discussion 3: Investigating factors related to enjoyment

Lack of association between our **in-game learning measures** and students' **enjoyment dimensions**

Discussion 3: Investigating factors related to enjoyment

Lack of association between our **in-game learning measures** and students' **enjoyment dimensions**.

→ The game environment does not impose any performance pressure on students **(Gee, 2003)**