An Introduction to the Measurement and Analysis of Video Game Interaction Data

Gregory K.W.K. Chung, Rebecca E. Buschang, Deirdre Kerr, and Danny Parks

An Introduction to the Measurement and Analysis of Video Game Interaction Data – NCME Training Session

San Francisco, CA – April 27, 2013
Odds and Ends

- Walk-ons
- Materials
  - Session schedule, 3 forms
  - UCLA evaluation form, NCME evaluation form
- iPads
- Work in groups – up to you
- Format – listen, do, reflect, discuss
- Informal - ask questions at any time
Overview of Session

• Introductions
  ✓ Greg Chung, Ph.D. – Assistant Director for Research Innovation
  ✓ Rebecca Buschang, Ph.D. – Senior Researcher
  ✓ Deirdre Kerr, M.A. – Graduate Student Researcher
  ✓ Danny Parks, M.S. – Lead Game Developer

• Assumptions
  ✓ Not a session on game design
  ✓ You have a game, you’re modifying a game, or you’re instrumenting a game
  ✓ You’re interested in measuring learning
## Schedule

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Topic</th>
<th>Duration (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Greg</td>
<td>Overview of session and introductory remarks</td>
<td>15</td>
</tr>
<tr>
<td>Rebecca</td>
<td>Extracting meaningful events and measures from gameplay</td>
<td>80</td>
</tr>
<tr>
<td>De</td>
<td>Approaches to the analyses of gameplay data</td>
<td>75</td>
</tr>
<tr>
<td>Danny</td>
<td>Rolling your own: Best practices from a game developer's perspective</td>
<td>30</td>
</tr>
<tr>
<td>Greg</td>
<td>Wrap up, fill out evaluations</td>
<td>15</td>
</tr>
</tbody>
</table>
Research Context

• Games, games, games
  
  ✓ Use of educational games to support teaching and learning is increasing at rapid rate, gaining acceptance by teachers, schools, and policymakers (Bienkowski, Feng, & Means, 2012)

  ✓ Use of games as a means to measure student learning (e.g., Baker, Chung, & Delacruz, 2012; Shaffer & Gee, 2012; Shute & Ke, in press)

  ✓ Growing research base hinting at the utility of games for learning (e.g., Tobias, Fletcher, Dai, & Wind, 2011)

• Data, data, data
Why Use Games to Measure Learning?

• Effortful performance on part of students

• Well designed game mechanics require use of targeted knowledge and processes
  ✓ Links behavior to cognition

• Well designed data logging captures meaningful player actions and context of action
  ✓ Able to derive useful measures
Why This Session?

• IES-funded 5-year R&D center: Center for Advanced Technology in Schools (CATS)

• Design Studies
  ✔ 24 studies total (10 experimental, 7 assessment), \( N = \sim 2500 \)
  ✔ Schools (grades 4-12), after-school program, private school
    ✤ 18 sites; 8 school districts; 4 states

• RCT (62 classrooms, \( N = \sim 1700 \) students)

• Why would you *not* focus on in-game process data?

• Bootstrap others new to the area
What to Measure
Common measurement strategy by non-learning types

Let’s build a game!
What should I measure?
Whatever is important!

Whatever the game is supposed to teach.
Whatever the game happens to do well.
Whatever is unique about the game.

Students liked the game better than regular instruction.
How to Measure
Common measurement strategy by non-learning types

Let’s use a game!

How do I measure “it”? It depends on “it.”

Didn’t the programmer do that?

Um ... how do I analyze this?

Students liked the game better than regular instruction.
Gameplay as a Data Source
Design properties that lead to $p < .001$

- Game mechanic requires use of target knowledge
- Game allows player to fail (i.e., commit errors)
- Game requires player to make decisions
- Player cannot progress without requisite knowledge
- Stage / level design useful [stages = 1 concept; levels = variation within concept]
  - Pigeon curve; inverse sawtooth wave; survival plot
- Rebecca’s section
Gameplay as a Data Source
Some really useful game logging properties

- Log behavior and not inferences
- Log cognitively meaningful events
- Data should be captured at the smallest usable grain size
- Logging should preserve uniqueness
- Data should be stored in a structured format
- De’s section
Extracting Meaningful Events and Measures from Gameplay

Rebecca Buschang
Next 80 minutes

- Introduction to the game
- Play the game
- Game as a measurement device
- Play the game
- Debrief
Learning Goals of Game

- The size of a rational number is relative to how one whole unit is defined.

Behr, Harel, Post, & Lesh, 2003; Fuson, 2003; Lamon, 1999
Learning Goals of Game

- Only identical units can be added to create a single numerical sum.

De Corte & Verschaffel, 1987; Kilpatrick et al., 2001; Mack, 1990

\[
\begin{align*}
1 + 2 + 3 &= \ ? \\
\frac{1}{4} + \frac{2}{4} + \frac{3}{4} &= \ ? \\
\frac{1}{4} + \frac{2}{3} + \frac{3}{5} &= \ ?
\end{align*}
\]
Level Sequence

- Whole units
- Fractions: Identifying correct denominator
- Adding fractions
- Create correct denominator
- Converting common denominator
Play game (15 minutes)

As you play, think about how the game could be a measurement device

(What would you measure in this game?)
Debrief: Game as Measurement Device

• In-game behaviors can be measured
  ✓ Deaths and successes
  ✓ Attempt information
  ✓ Common errors
  ✓ Math behaviors specific to certain levels
Play game (20 minutes)

- Think about the game behaviors, and how the game behaviors are linked to cognition and math understanding.
- Write down what you think are the important behaviors to capture.
Debrief

• What game behaviors are important to capture?
• How are the game behaviors linked to cognition? Math understanding?
• What are the most important behaviors to capture?
Break (20 min)
Approaches to the Analyses of Gameplay Data

Deirdre Kerr
Next 75 minutes

- Logging the game
- Practice logging
- Extracting summary statistics
- Evidence identification
- Interpret cluster analysis
- Using cluster analysis results
- Plan an analysis
- Debrief
What Does Logging Do?

- Records what the student did in the game
- Saves to permanent file
- Saves in structured format
- Analyzed to make inferences about student:
  - Performance
  - Understanding
  - Process
Why Focus on Logging?

• Can’t measure what you don’t record
  ✓ What research questions should the data address?
  ✓ What level of granularity do you need?

• Can’t record everything*
  ✓ What should you record?
  ✓ What inferences should you make while recording?
Logging Should Record...

- Actions not answers
- Only “important” actions
- Context information
- General and specific information
- In a structured format
The Other Side of the Fence

• Create an algorithm (or many algorithms) that calculate the probability that the student understands each concept

• Update the probability as they play

• Print out the final probability and use it to make inferences about the student
Our Side of the Fence

- Allows for summary statistics
- Allows differentiation between the same actions in different contexts
- Allows differentiation between user actions and programmed game actions
- Allows for knowledge discovery through various forms of data mining
## Example: Logging

<table>
<thead>
<tr>
<th>ID</th>
<th>Game Time</th>
<th>Data Code</th>
<th>Data Description</th>
<th>Data 01</th>
<th>Data 02</th>
<th>Data 03</th>
</tr>
</thead>
<tbody>
<tr>
<td>1115</td>
<td>3044.927</td>
<td>2050</td>
<td>Scrolled rope from [initial value] to [resulting value]</td>
<td>1/1</td>
<td>3/3</td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>3051.117</td>
<td>3000</td>
<td>selected coil of [coil value]</td>
<td>1/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>3054.667</td>
<td>3010</td>
<td>added fraction at [position]: added [value] to yield [resulting value]</td>
<td>1/0</td>
<td>1/3</td>
<td>1/3</td>
</tr>
<tr>
<td>1115</td>
<td>3058.443</td>
<td>3000</td>
<td>selected coil of [coil value]</td>
<td>1/3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>3054.667</td>
<td>3010</td>
<td>added fraction at [position]: added [value] to yield [resulting value]</td>
<td>1/0</td>
<td>1/3</td>
<td>2/3</td>
</tr>
<tr>
<td>1115</td>
<td>3088.886</td>
<td>3020</td>
<td>Submitted answer: clicked Go on [stage] – [level]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>3097.562</td>
<td>3021</td>
<td>Moved: [direction] from [position] length [value]</td>
<td>Right</td>
<td>1/0</td>
<td>2/3</td>
</tr>
<tr>
<td>1115</td>
<td>3106.224</td>
<td>4020</td>
<td>Received feedback: [type] consisting of [text]</td>
<td>Success</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1115</td>
<td>3108.491</td>
<td>5000</td>
<td>Advanced to next level: [stage] – [level]</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Log an Attempt
(10 minutes)

Data Description:
text description of the action with brackets around variables
  e.g. add [value] to [position]

Data_01 to Data_04:
values for each variable in the description

Context:
game state information
Extracting Summary Statistics

- Calculate last level reached
- Determine mean number of attempts
- Sum up individual data codes
  - ✅ Clicked Help (4010)
  - ✅ Hit Reset (3030)
  - ✅ Added ropes to a sign (3010)
  - ✅ Scrolled through ropes (2050)
Evidence Identification

• How do you know what they did?
  ✓ Read the data
  ✓ Report case studies

• What if there is too much data to look through by hand?
  ✓ Have a computer read it (data mining)
  ✓ Name identified groups
  ✓ Report/analyze identified groups
Cluster Analysis

- Uses unique actions as input
- Clusters/groups actions that occur “together”
  - By the same students
  - In the same level
  - In the same attempt
- Outputs “strategies”
  - Named action groups
  - Solution types and error types (game and math)
Fuzzy Cluster Analysis

• One action can be in more than one cluster
  ✓ Branching strategies
  ✓ Cluster analysis outputs “belongingness” probability

• No indicators of time
  ✓ Absolute or relative

• Each action only appears once in a given cluster
  ✓ Regardless of how many times it occurred
Interpret Cluster Analysis
(10 minutes)

Name each cluster:
Come up with a name for the strategy that would explain that set of actions

What other strategies might students have used?
Are there any other errors you would expect?

What actions would identify those strategies?
What evidence would you find in student gameplay to indicate that was happening?
Using Cluster Analysis Results

• What strategies are most common?
• Is strategy use predicted by demographic variables?
• How does strategy use affect performance?
• How does strategy use affect learning?
• Is strategy use related to game design elements?
Plan Advanced Analyses (10 minutes)

Create a research question:
What do you want to know about strategy use or what can strategy use tell you?

Determine how you will measure the strategies:
Mean number of attempts in each strategy?
Used in first attempt? Count of attempts?

Determine your methodology:
Regression? Structural equation modeling?
Bayesian networks? IRT?
Debrief
Break (5 min)
A Developer’s Perspective
Danny Parks
Next 30 minutes

- Defining “Important”
- Pitching to programmers
- Use cases for data
- Formatting
- Local or web deployment
- Scaling up
- Open issues
What Is Important?

Lossless Playback

http://www.nammfoundation.org/system/files/happymusic.jpg
http://rogueamoeba.com/fission/images/screenshots/FS_sshot5.png
Pitching to Programmers

Avoiding push back

http://molovinskyonallentown.blogspot.com/2012/02/association-of-realtors-used-car.html
How Should I Write My Logs?

Who Is Using the Data?

Researchers

Software

Developers
Different Priorities

- Getting it to work
- Knowing what it means
- Crunchability

http://farm3.static.flickr.com/2777/4112883676_6f175d74af.jpg
Options

• Multiple logs
  ✓ Potentially a mess
  ✓ Each user gets only the info they want

• Parsers to clean or reformat data
  ✓ Only one log file
  ✓ Requires using another piece of software before data is useful
  ✓ High probability of lost data
The Answer

- Document Everything!
  - Why did we do it like this?
  - What does this mean?
  - Where can I find this?
  - What were we thinking?

http://www.thelandscapeoflearning.com/2012/10/questions-have-always-been-answer.html
Readability and Size
Adding readability makes files bigger

Who Cares?
Just log it all!

http://bobgroves.com/have-you-read-these-6-books-5-out-of-6-dentists-say-you-should/pile-of-books/
PBS: 1+ gigs/game/day

http://cdn.mhpbooks.com/uploads/2012/05/Curious_Georg.jpg
Local or Web?
A simple question with big implications

Let’s Talk About the Internet…

Go Big, or Go Home?

<table>
<thead>
<tr>
<th>Pros</th>
<th>Local</th>
<th>Web</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Controlled testing environment</td>
<td>• Collecting the data</td>
</tr>
<tr>
<td></td>
<td>• Files are all on the local machine</td>
<td>• Distributing the game</td>
</tr>
<tr>
<td></td>
<td>• No need for Internet connection</td>
<td>• Files are all on the server</td>
</tr>
<tr>
<td>Cons</td>
<td></td>
<td>• Internet connectivity</td>
</tr>
<tr>
<td></td>
<td>• Distributing the game</td>
<td>• Browser compatibility</td>
</tr>
<tr>
<td></td>
<td>• Collecting the data</td>
<td>• Users</td>
</tr>
</tbody>
</table>
Other Considerations

• How much?
  ✤ *Do you really need all of those events?*

• Where do you log?
  ✤ *Local machine, online database, both?*

• When?
  ✤ *After every event, every level, every day, ever?*
Going Big

- Database setup
  - Sorting and retrieving data
  - Multiple games at once
  - Multiple conditions per study

www.grtcorp.com/sites/grtcorp.com/files/going_big.jpg
Going Big (cont.)

• Balance traffic and processing
  ✓ Package size and number of packages
  ✓ Sorting and storing
  ✓ Real-time analysis

Fail Gracefully
How do we turn it off!?!
Who Are These People?

- How do we know who generated this data?

ID Strategies

• User IDs and attribution
  ✓ Single Sign On
  ✓ “Unique” IDs (local!)
  ✓ Logging in later
Unknown Unknowns
Logging things you can’t know

✓ Correct solutions
✓ Player intentions

Open Problems
Things I don’t know how to do

✓ Real time simulation states
✓ Co-op attribution

http://www.johnst.com/data/blog/7/227_522.jpg
Summary

- Where is the data coming from?
- How is it going to be used?