Developing Games to Measure Personality—It's Not All Fun and Games!

Andrea Sinclair
asinclair@humrro.org

Interdisciplinary Workshop on Game-based Assessment
August 22, 2019
Minneapolis, MN

Context

- Personality measures are popular tools for hiring decisions
  - Incremental r_{xy} over *g* for job performance
  - Small subgroup differences (race/ethnicity)

- Self-report personality measures assume . . .
  - personality is best measured by asking people how they tend to behave, and
  - people can accurately report how they behave

- However, self-report personality measures are . . .
  - susceptible to faking, and . . .
    - faking is difficult to prevent

Potential Remedy

- Game-like performance tests (GLPTs)
  - A measure (a) that requires people to perform a task in a video game format, and (b) for which people's behaviors in the game relate to their personalities
- GLPTs may reduce faking because people are playing a game instead of focusing on the best way to answer questions about their personality
- Extant research on the relations between performance test scores and personality scores is sparse and non-definitive

Objectives

- Explore the creation and construct validation of video GLPTs
- Assess personality unobtrusively via GLPTs
- Develop GLPTs that are less susceptible to faking and impression management
- Develop simple, yet engaging games
  - Don’t want to measure . . .
    - cognitive ability
    - game-playing experience
    - psychomotor ability
- Manipulate situational strength within the GLPTs

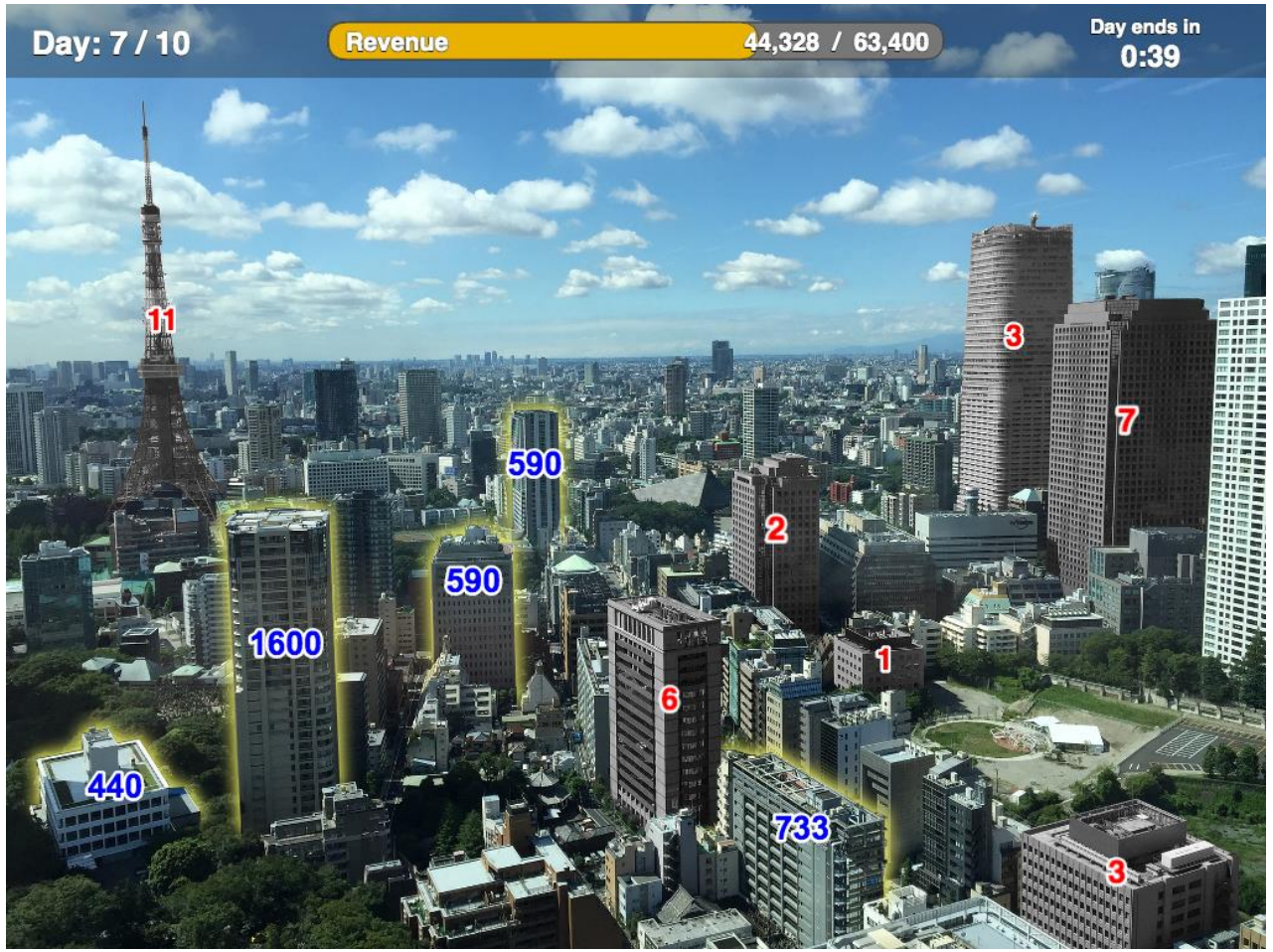
Project Team

- Practitioners (HumRRO team)
  - Andrea Sinclair
  - Rod McCloy
  - Justin Purl
  - Amanda Koch
- Academics
  - Richard Landers
  - Fred Oswald
  - Reeshad Dalal
- Game-developer
  - Insystech (Mathew Varghese, Ramji Venkatachari)

The Games!

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION

The Game: Click Town



Click Town (Achievement-striving)

- Focal Behavior
  - Will the player stay to collect additional revenue or exit the town as soon as possible?

- Scoring
  - 217 scorable events identified (so far)
  - 6 categories:
    a) clicks,
    b) clicks that produce revenue,
    c) clicks that do not produce revenue,
    d) the speed at which productive clicks are made,
    e) the speed at which unproductive clicks are made, and
    f) the overall revenue collected.

Situational Strength Features

- Instructions inform players to pay attention to the . . .
  - timer and
  - to their progress in meeting the revenue goal
- Level timer (which starts flashing at 10 seconds left)
- Progress in meeting revenue goal displayed
- A message is displayed when the Next Day button appears

Situational Strength Features

Level timer

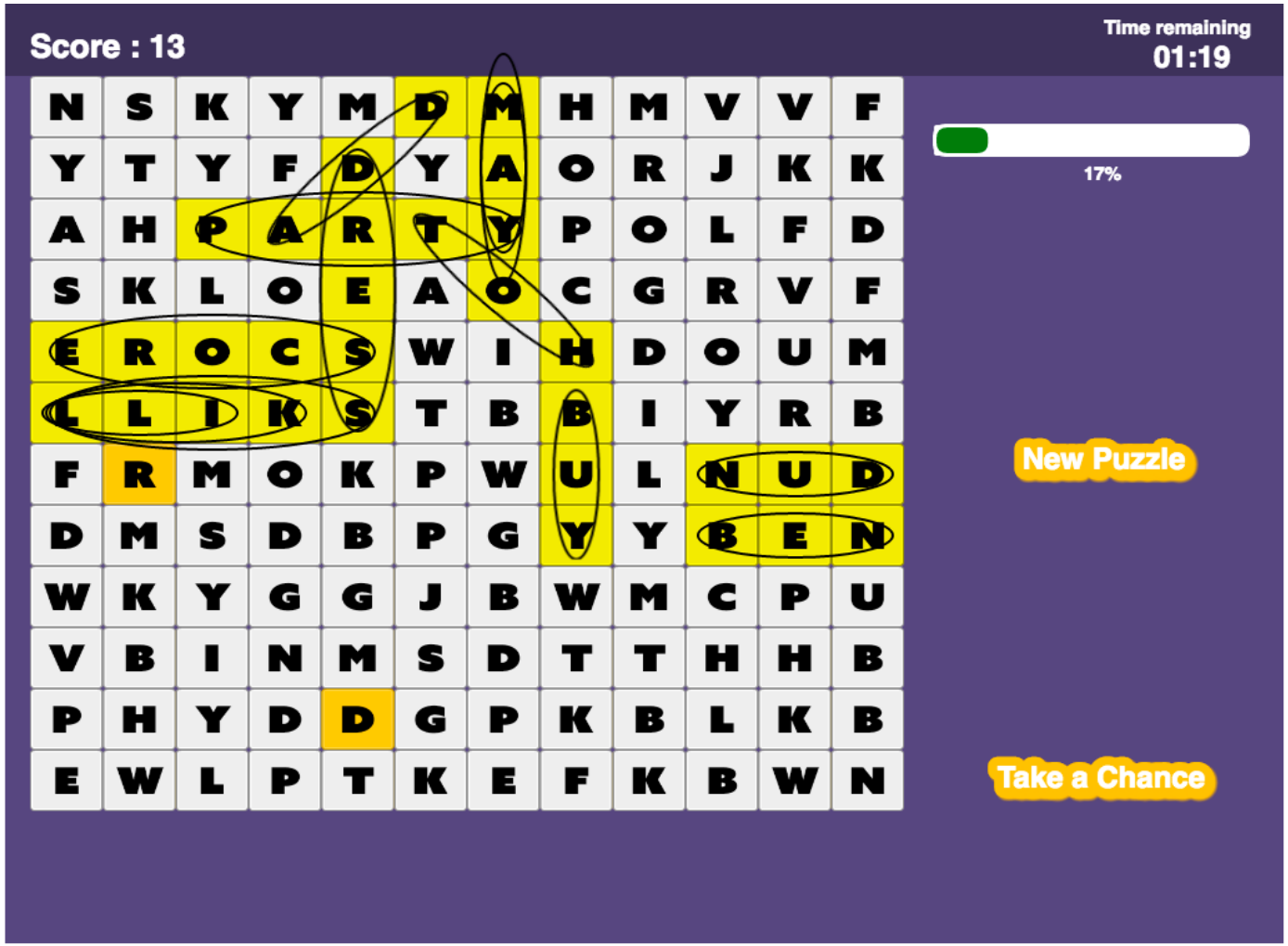
- *Starts flashing when 10 seconds left*

Progress in meeting revenue goal displayed

Situational Strength Features

A message is displayed when the Next Day button appears

The Other Game: Word Find



Challenges

- Narrowing down the personality traits to measure
  - Achievement-striving: going “above and beyond”
  - Perseverance: pursuing a task to completion; persisting despite adversity
- Working within budget and timelines
  - Originally planned to develop four games (two for each personality facet)
- Communication between Researchers and Game Developers
  - Highlighted difference between developing a GBA and gamifying an assessment

- Many, many decisions to make
  - e.g, color of buildings, style of buildings, transition time between screens, size of numbers, location of numbers, etc.
  - The many decisions are compounded by the fact that have to think about how the decision might interact with measurement of the personality trait of interest
- Manipulating situational strength
  - Activating trait-relevant cues while maintaining low situational strength
- Creating the database and determining how to “score” captured behavior
  - Around 1.2 million data points in just one of the games

Results

- For the most part, the correlations we observed were weak but positive, indicating something more systematic than a random event.
  - The sum of the correlations with achievement striving across the 10 scorable events for Click Town was .32 in the low SS condition.
  - The sum of the same 10 correlations in the high SS condition was just .03.
- Thus, although there is only a weak relation, there seems to be a relation nonetheless.
- Controlling for cognitive ability and psychomotor ability had little impact on the observed correlations, thus indicating that our attempts to design the games to be simple (so as to avoid undue assessment of these two abilities) was largely successful.

Summary & Future Directions

- Measuring personality in a game is hard!
  - Nonetheless, we think there's promise.

- Future Directions
  - More refined modeling techniques (e.g., machine learning regression models)
  - Produced millions of data points
  - Each record contains lots of information (timestamps for Building 1 timer started, Building 1 clicked, Building 1 timer expired, Building 1 revenue collected, etc.)
  - Many ways to combine the data into scores that describe a player's game behavior
    - Composites of scores
    - Patterns in scores

Thank you!

HumRRO
HUMAN RESOURCES RESEARCH ORGANIZATION