**Analysis protocol**

# Learning network complexity through gameplay: An experimental study of data game learning outcomes

2021

*Authors:*

Chris Moreno-Stokoe, Claire Haworth & Oliver Davis

#### Linked document statement

This document is part of a GitHub repository and, unless otherwise provided, refers to documents within this repository:

|  |  |
| --- | --- |
| *Linked documents*  <https://github.com/CMorenoStokoe/gsa> | |
| **Description** | **Location** |
| Intended Learning Outcomes | ./reference/ILOs v1b.xlsx |
| Assessment Questions and Answers | ./reference/Questions and answers v1b.docx |

# Background

# Brief introduction

Games are a good medium for exploring complexity because data can be accurately modelled in games and players can be directed and motivated to engage with the data more than they otherwise would. The literature on game-based learning demonstrates that games can enhance the learner’s experience and this can translate to improvements in Intended Learning Outcomes. For example, games can provide concrete experiences and active experimentation (Ricardi & DePaulis, 2014) and constructively aligning the learning and assessment modalities to the degree that in-situ stealth assessments are possible (Susi, 2007). Simulation games are particularly popular and are in widespread use for surgery, military, industrial applications and even in the policy making decision process (Duke and Geurts, 2004; Bilson, Bekebrede & Mayer, 2010).

This chapter extends the previous literature on game-based learning to complex networks of the relationships between factors in public health (e.g., BMI, smoking and wellbeing). In addition, previous studies on simulation do not clearly delineate between gameplay and simulation experiences but we will experimentally separate the gameplay from an interactive visualisation simulation. We experimentally investigate whether gamifying an interactive visualisation simulation improves learning outcomes about a public health data network.

Bloom’s Taxonomy of Learning (e.g., Anderson & Bloom, 2001) tells us that there are different levels of understanding. These are often visualised as a pyramid, with the lowest and easiest levels being the ability to remember the information; interpret, exemplify, and summarise it, and the higher and harder levels concerning critical evaluation and ability to use the information. It therefore follows that the best test of high-level understanding of complex public health dataset is to ask learners to use this information to design public health interventions. Learners who are capable of this will demonstrate a high degree of understanding.

# Aims

##### Research question

Does gamification improve student’s experience and outcomes of learning the effects between traits in a public health intervention simulator?

##### Objectives

We experimentally compared a gamified and non-gamified interactive visualisation in order to investigate the following hypotheses:

* Gameplay will alter the learning experience
  + It will provide structure for engaging with learning software
  + It will motivate learners to engage with the learning software
  + It will facilitate better learning of public health network data

Additionally, we explored how specific game mechanisms modify the learner experience

# Methods

## Design & Materials

An interactive visualisation was developed for network public health data (network MR estimates). This was developed into a game with minimal modifications so that it could serve as a control in a ‘gamified’ vs ‘non-gamified’ experimental comparison. A document fully describing the learning software (interactive visualisation and game) will be available following write-up of chapters 3 (MiRANA) and 4 (development of this game). A document fully describing of the public health data will also be available following final write-up of chapters 2 (MR study) and 3 (MiRANA).

Details of the online form materials used for conducting the study are given below combined with the procedure.

Ethics approval was obtained from the University of Bristol Psychological Science School Research Ethics Committee (ID: 111083).

## Procedure

Participants were directed to complete an online form. This form contained four sections:

1. **Participant information and consent**

Participants were given information relating to the study and required to give their consent before continuing with the study.

1. **Assignment to experimental condition**

Participants were randomly assigned to either an experimental or control condition. In the control condition, participants were given an interactive visualisation without gamification (<https://www.morenostok.io/mendel/interactiveVisualisation.html>). In the experimental condition, participants were given a gamified form of this visualisation (<https://www.morenostok.io/mendel/game.html>). These were assigned as the participant’s learning software for the experiment, and were prompted to access them by URL link.

On the same page of the online form, the number of mouse clicks and time duration spent on page were measured, and participants were asked if they had taken a break while on that page (and, if so, for how long). Individuals were prompted to “Spend as long as you like until you feel like you have a good enough understanding of the relationships between public health traits and then return to this questionnaire for an open-book learning assessment”.

1. **Self-report feedback**

Participants were then asked about their experience with the software. Four questions were asked:

* Please indicate your agreement with the following statements using 7-point Likert scales (strongly agree - strongly disagree):
  + I felt motivated to make interventions
  + I felt guided to interact with the simulation in a specific way
* Please drag the experiences below to the appropriate box depending on whether you experienced them or not (based on Playful Experiences Framework: Lucero et al., 2013):
  + Captivation - Forgetting one’s surroundings
  + Challenge - Testing abilities in a demanding task
  + Competition - Contest with oneself or an opponent
  + Completion - Finishing a major task, closure
  + Discovery - Finding something new or unknown
  + Progression - Earning momentum and achievement
  + Exploration - Investigating an object or situation
  + Fantasy - An imagined experience
  + Humor - Fun, joy, amusement, jokes, gags
  + Nurture - Taking care of oneself or others
  + Relaxation - Relief from bodily or mental work
  + Sensation - Excitement by stimulating senses such as sights or sounds
* A free-form answer question:
  + In your own words please describe your experience with the software. For example, did you have a strategy? Did anything prevent you from achieving what you wanted? Did you find any effects memorable? Did you have any opinions about the presentation?

1. **Learning assessment**

For the final section involved a learning assessment intended to measure learning of the relationship between traits in a public health dataset. The dataset the game was based on, and its use in the game, will be fully detailed in the document following write-up of chapter 4 (development of the game). In this section of the online form, participants were given an introduction to direct and indirect effects, an essential term which participants required a basic understanding of to answer the following questions. Participants were also prompted to open, and keep open for reference, a visualisation of the data via URL link (<https://www.morenostok.io/mendel/visualisation.html>). The assessment was an open book test in order to test better player’s implicit learning rather than rote learning, or memorisation of the traits in analysis.

The test took participants about half an hour to complete and involved five progressively more difficult sections which tested seven areas of competency (see table 1). Each of these areas of competency are matched with specific intended learning outcomes and corresponding assessable competencies which individuals should be able to perform (listed in full in the Intended Learning Outcomes linked document).

**Table 1.** Areas of competency.

1. Ability to read information about nodes and edges in the network visualisation
2. Understanding of direct effects: Infer the direct effects of interventions which increase the prevalence of a trait
3. Understanding of network properties: Ability to make general inferences about the network
4. Understanding of interactions: Critically analyse interaction effects between multiple interventions which increase the prevalence of different traits
5. Ability to negate effects: Infer the direct effects of interventions which decrease the prevalence of a trait
6. Understanding of indirect effects: Infer the indirect effects of interventions which increase the prevalence of a trait
7. Working understanding of complex network effects: Ability to design solutions to public health problems involving multiple interventions which have a mix of direct and indirect effects, which both increase and decrease traits

The learning assessment (Q n=29) has two modalities; the first 25 questions are delivered as a multiple-choice questionnaire. The last 4 questions are a free-form intervention design exercise. All questions and scoring are detailed in the Assessment Questions and Answers linked document.

# Analysis plan

# Scoring

The two (multiple-choice and free-form) modalities are scored differently. If it appears that both tests load onto each other (i.e., strongly correlate with each other), these scores will be merged into a single learning assessment score. Conversely, if there is evidence that they measure different constructs (i.e., high scorers for one test are not high scorers in the other), then they will be treat as two separate tests and two learning scores will be devised for each participant.

**Multiple-choice assessment**

The first 25 (multiple-choice) questions were scored simply according to the number of correct answers (for scoring per question see list of questions). Scores will be adjusted by subtracting the score obtainable by random guessing.

**Free-form assessment**

The last 4 (free-form) questions were intervention design scenarios. For these questions, individuals were scored by how much their intervention achieved the optimal results. In order to discover the optimal results, network MR will be performed to calculate all of the effects of an intervention on each trait in analysis. This is the same calculation the game uses to score players, and is based on MiRANA’s JavaScript network MR model, and will be detailed in full as part of chapters 3 (MiRANA) and 4 (developing the game).

Each individual effect of an intervention will be scored according to whether the change was positive/negative and whether the trait was good/bad/neutral (see Table 1 for traits valences). For example, reducing smoking would give a high score, but reducing wellbeing would not.

**Table 1.** Categorisation of trait valence

|  |  |  |
| --- | --- | --- |
| **Bad** | **Neutral** | **Good** |
| Depression | Coffee intake | Exercise |
| Worry | BMI | Sleep |
| Loneliness | Eveningness | Wellbeing |
| Insomnia |  | Education |
| Neuroticism |  | Intelligence |
| Alcohol |  | Socialising |
| Smoking |  |  |
| Heart disease |  |  |
| Diabetes |  |  |

This will be performed for every possible combination of traits for intervention, and from this an optimal intervention will be selected. Individuals will receive scores based on how effective their intervention was compared to the optimum. These are standardised to each question so that the maximum us 100% and minimum is 0%. For example, if an individual scored 50% on a question, then they would have produced an intervention which was half as effective as the optimal intervention.

Interventions will be scored in terms of their primary and side-effects.

Primary effect score

The primary effect of an intervention describes how it changed the prevalence of the objective trait (e.g., for a question asking you to reduce smoking the objective trait would be smoking). The primary effect of an intervention is scored simply as a percentage of the maximum possible good effect:

Side effect score

The side-effect score for an intervention describes its overall effects on other traits in the network besides the objective trait (e.g., accidentally increasing heart disease as a side-effect while trying to reduce smoking). Interventions which increase good traits and reduce bad traits will receive high scores.

**Timing**

Participants were timed and their number of mouse clicks recorded while on various pages throughout the online form. It is not clear how reliable mouse click tracking is so this will not be analysed further. Combined with self-report whether they took breaks while these pages were open on their computer, this will be used to measure how long individuals spent on these pages:

* Page where individuals are assigned to a condition and directed to learning software

The first will be used as a measure of time spent engaging with the learning software.

* Each page in sections 1-7 of the learning assessment

These will be summed as a measure of how difficult an individual found the learning assessment.

# Hypothesis testing

Hypotheses will be tested using T-tests with an accepted alpha level of 0.05. A hypothesis will be rejected if all of its tests return non-significant results. The three hypotheses will be tested as follows:

**Gameplay will alter the learning experience…**

**It will provide structure for engaging with learning software**

* Compare participants Likert responses to the item ‘I felt directed to use the software…’ between gamified/non-gamified groups

**It will motivate learners to engage with the learning software**

* + - Compare participants Likert responses to the item ‘I felt motivated to use the software…’ between gamified/non-gamified groups
    - Compare the time participants used the gamified/non-gamified software
    - Explore responses given in the open-form text box about their experience with gamified/non-gamified software

**It will facilitate better learning of public health network data**

* Compare participants scores between gamified/non-gamified groups
* Compare the time it took participants to complete the assessment between gamified/non-gamified groups

# Planned exploratory analysis

How do specific game mechanisms modify the learner experience?

* + - Comparing the experiences participants reported feeling (PLEX) to see if the game led to feelings of ‘competition’ or ‘thrill’ for example
    - Explore responses given in the open-form text box about their experience