

# An empirical study of the impact of Systems Thinking and Simulation on Sustainability Education

Caroline Green\*, Owen Molloy and Jim Duggan

School of Computer Science, National University of Ireland Galway, Galway, Ireland

caroline.green@nuigalway.ie; owen.molloy@nuigalway.ie; james.duggan@nuigalway.ie

\* Author to whom correspondence should be addressed. ORCID ID:

<https://orcid.org/0000-0003-4811-5192>.

**Abstract:** Education for Sustainable Development (ESD) is considered vital to the success of the United Nations' Sustainable Development Goals. Systems Thinking has been identified as a core competency necessary to incorporate into ESD. Systems Thinking orientated ESD learning tools, established methods of assessment of sustainability skills, and studies to demonstrate effectiveness of such learning tools, are all lacking.

There is a wealth of experience in the System Dynamics field regarding the application of Systems Thinking and simulation to environmental problems, sustainability and systems education. Many System Dynamicists regard simulation as essential for teaching Systems Thinking. The substantial body of research into the design of effective simulation-based learning environments (SBLEs) can also inform ESD initiatives.

This research describes a randomised controlled trial (n=106) to investigate whether an online sustainability learning tool that incorporates Systems Thinking and System Dynamics simulation increases understanding of a specific problem and supports transfer of knowledge to a second problem with a similar systemic structure. The effects of Systems Thinking and simulation were tested separately and in combination. The learning tool was designed for a single online learning session. Simulation was found to increase ESD learning outcomes significantly, and also to support transfer of skills, although less significantly.

**Keywords:** Education for Sustainable Development, Systems Thinking, System Dynamics, simulation, transfer of skills, effectiveness, randomised controlled trial, RCT, factorial study, ANOVA

---

## Introduction

Sustainability has become an increasingly important topic of discussion over the last several decades as the harmful long-term consequences of unsustainable human activities and lifestyles have become ever more apparent. However, the concept of sustainability is complex and is often used in an imprecise way [1]. Since the 1980s the UN has been instrumental in developing the related concepts of sustainability, sustainable development and sustainability education. It founded The World Commission on Environment and Development (WCED) in 1980 which was responsible for the influential 1987 Brundtland Report [2]. The definition of sustainability in that report is the one most frequently quoted, namely that 'Sustainable development is



development that meets the needs of the present without compromising the ability of future generations to meet their own needs.'

The UN also led efforts to formulate concrete targets for action towards sustainability. In 2000 the UN defined the 8 Millennium Development Goals (MDGs) for 2015, of which goal 7 was 'To ensure environmental sustainability'. The MDGs were developed further in the 17 Sustainable Development Goals (SDGs), set in 2015 and to be achieved by 2030, and adopted by all 193 United Nations member states.

The UN adopted the Decade of Education for Sustainable Development (DESD) from 2005 to 2014. Education for Sustainable Development (ESD) is explicitly recognized in the SDGs as part of Target 4.7 of the SDG on education. It is seen as 'crucial for the achievement of sustainable development' [3] (p. 63). The Council of the European Union sees ESD as 'essential for the achievement of a sustainable society and is therefore desirable at all levels of formal education and training, as well as in non-formal and informal learning'<sup>1</sup>. Thus ESD is seen as a form of lifelong learning, and necessary for all citizens.

Sustainability education seeks to address the considerable challenge of training learners not only to solve or understand existing complex problems but also to equip them with skills that they can transfer to new problems as they arise. In the last few years there has been an urgent call for innovative sustainability pedagogies [4] (p. 58).

O'Flaherty and Liddy provide a useful summary of approaches so far taken to ESD, including blended learning, drama, simulation exercises, multi-media, problem-based learning and discussion forums [5]. They describe methodological and pedagogical questions that remain open and highlight the need for assessment frameworks and formal trials for evaluating the effectiveness of different approaches to ESD.

#### *The need for Systems Thinking in Sustainability Education*

Sustainability education is an emerging field. In her review, Maria Hofman-Bergholm explores reasons for problems with its implementation [6]. She draws out the commonalities between the literatures on Sustainability Education and Systems Thinking, in that both require critical thinking, real-world complex problem-solving skills and action. She finds that Systems Thinking is required to comprehend the intricate connections in sustainable development [7] (p. 27). Complex reasoning skills must be taught, as they are not inherent. Humans have well-known limitations in cognitive ability to reason about complex systems that must be overcome [8] (p. 599).

Similar observations have been made in the related field of Ocean Literacy [9]. A pilot study we conducted to investigate the effectiveness of a systems-orientated online Ocean Literacy learning tool gave promising results [10].

According to Frisk and Larson, sustainability education will only be effective if it incorporates Systems Thinking, long-term thinking, collaboration and engagement, and action-orientation [11]. Sustainability, they say, is fundamentally a call to action, and Sustainability Education therefore requires experiential, practical and flexible learning methods.

According to Wiek *et al.*, 'Sustainability education should enable students to analyse and solve sustainability problems' [12] (p. 204). This

<sup>1</sup> Council of the European Union. Council conclusions on education for sustainable development. Viewed 9 June 2021. <[https://www.consilium.europa.eu/uedocs/cms\\_data/docs/pressdata/en/educ/117855.pdf](https://www.consilium.europa.eu/uedocs/cms_data/docs/pressdata/en/educ/117855.pdf)>

requires a particular set of interlinked and interdependent key competencies. Wiek et al. review the literature and identify five key competencies, the first being Systems Thinking competence (the others are anticipatory, normative, strategic and interpersonal competence).

According to Soderquist and Overakker, the discipline of Systems Thinking provides a process, set of thinking skills and 'technologies' that can improve the systemic understanding that is required for sustainability education [13]. These include stock and flow mapping, computer simulation, and simulation-based learning environments. They claim that simulation-based learning environments build mental simulation capacity, if they are designed carefully.

Cavana and Forgie describe a number of well-established systems education programs and review teaching approaches for Sustainability Education [14]. They explore the strong links between systems approaches and sustainability goals, illustrating that the two are so entwined as to be inseparable. They describe the need for, and the lack of, simulation-based learning environments for Systems Thinking orientated sustainability education.

#### *Relevant work in the field of Systems Dynamics*

A substantial body of knowledge focused on modelling and simulation of complex human-environmental systems has accumulated in the field of System Dynamics since the 1970s. This is potentially useful for informing efforts to develop effective, innovative Systems orientated sustainability education tools.

System dynamics modelling was first used to address sustainability in Jay Forrester's 'World3' model, which formed the basis for the influential book, 'Limits to Growth' [15]. There have been many subsequent examples, from environmental models [16,17], models for water supply, waste management, air quality, land use [18], fisheries [19,20], climate change [21], models of social and economic development [22], reindeer pasture management [23] and many more.

Furthermore, the System Dynamics community has identified education as a priority for a long time. The Creative Learning Exchange was founded in 1991 by Jay Forrester 'to encourage the development of systems citizens who use systems thinking and system dynamics to meet the interconnected challenges that face them at personal, community, and global levels'<sup>2</sup>. They provide resources representing three decades of experience of teaching Systems Thinking and System Dynamics for real-world problem-solving to school children [24].

System Dynamics simulation has frequently been employed for the purpose of environmental education. There are flight simulators for sustainability [25] and simulation-based learning environments to teach sustainability [26,27]. System Dynamics models and simulations have also been used to try to explain why renewable resources are so often over-utilised; this is because of faulty reasoning and systematic misperceptions of the dynamics of complex systems [23]. Simulation has been shown to improve understanding and performance in a natural resource management task [28]. Simulation can serve effectively as the 'problem' in problem-based learning [29], and as an experiential activity it can both increase retention and have a stronger influence on behaviour than declarative learning [11] (p. 11).

---

<sup>2</sup> About The Creative Learning Exchange. Viewed 9 November 2021. <<http://www.clexchange.org/cle/about.asp>>

There is debate about whether simulation based on stock and flow models is an essential, advanced part of Systems Thinking or an extension of it [30]. According to Richmond, 'System thinkers use diagramming languages to visually depict the feedback structures of... systems. They then use simulation to play out the associated dynamics' [31]. Because simulation is seen as an essential by some, but not by all, systems thinkers, in our study the effect of adding Systems Thinking and simulation was evaluated separately and in combination.

System dynamics scholars have also long been interested in the transfer of insights between management situations that share common structural characteristics, going back to Forrester [32] (p. 355). According to Sterman, perhaps counterintuitively given the immensely rich and varied complex systems in the world around us, 'most dynamics are instances of a fairly small number of distinct patterns of behaviour' [8] (p. 108). Senge describes Systems Archetypes as 'nature's templates' [33] (p. 92). They reveal an elegant simplicity underlying complex issues. Mastering them represents putting Systems thinking into practice. Indeed, Richmond includes what he calls 'generic thinking' in his list of eight critical Systems Thinking skills [34]. Once an archetype is identified, 'it will always suggest areas of high- and low-leverage change.' For this reason, Kim views archetypes as diagnostic tools [35-37].

Because of the importance of Systems Archetypes to many systems thinkers, and because of the potential benefits of their use to sustainability education, their effect on transfer of sustainability skills is explored in this study. The choice of two sustainability problems that share a common systems archetype was made to test the hypothesis that learners can recognise similar patterns in different contexts, and therefore transfer their learning. If successful, this approach would make a strong case for a patterns-based approach to sustainability education, which would build systems and environmental literacy, obviating the need to teach each sustainability challenge in a piecemeal fashion.

#### *Design considerations for Simulation-Based Learning Environments*

The field of sustainability education can also benefit from the accumulated body of knowledge relating to design aspects and best practice in the development of simulation-based learning environments (SBLEs).

Landriscina advises that learners need guidance with simulations in the form of explanations, background information, tasks to perform, hints and feedback [38]. Kopainsky and Sawicka [28] (p. 143) cite Yasarcan [39] who holds that a 'gradual-increase-in-complexity approach helps improve performance in an inventory management simulation game'. In their critical review of 61 studies to evaluate effectiveness of simulations used for science instruction, Smetana and Bell report that 'simulations used in isolation were found to be ineffective', and that they should encourage reflection and promote cognitive dissonance, meaning that learners confront their erroneous assumptions and reconstruct their beliefs [40]. Cannon-Bowers and Bowers identified the importance of using case studies as a context for instruction and setting goals for the learner [41].

Ghaffarzadegan, *et al.* [42] argue that simulations based on small System Dynamics models offer advantages for learning in a public policy context. By small models they mean 'models that consist of a few significant stocks and at most seven or eight major feedback loops'. These

small models can 'yield accessible, insightful lessons for policy making' without overwhelming participants with too much detail.

There are two main approaches to simulation-based learning, learning by building a simulation, or by using an existing one. Reimann and Thompson assert that while learning by modelling may result in better long term learning outcomes, positive results have also been found in studies examining the effect of learning with pre-built models [43] (p. 115). Gobert and Buckley concur [44], stating that learners can gain more insight from building models, but considerable time and skills are required. If this is not feasible, manipulation of an existing simulation offers an alternative. The approach can vary from the simplest, where learners can change a few variable values and see the consequences of their decisions on graphs, to the more complex, where learners can restructure the model. Reimann and Thompson believe that, given the greater amount of time needed to train students to use modelling software, and for them to produce a working model, 'learning with pre-built models may be a more realistic option in an environmental education context'.

The ESD learning tool developed for this study was designed in line with these general guidelines. It was designed for a single online learning session and therefore interaction with the simulation model was limited to manipulation of a few key variables.

#### *Aims of the study*

Summarising the themes identified in the reviewed literature, the following research areas were identified and motivated the work described in this paper:

1. There is a need for Systems-based sustainability learning tools that can be shown to increase the effectiveness of sustainability education.
2. It would be useful to evaluate the effect of Systems Thinking (theory, tools and techniques) separately from that of interactive simulation, so that the effect of each factor on learning outcomes, and their combined effect, can be compared.
3. If Systems Thinking can facilitate recognition of similar systemic structures in different sustainability problems, this could make a useful contribution to the development of transferable sustainability skills.
4. Formal trials to evaluate the effectiveness of approaches to ESD, specifically a Systems Thinking approach, are needed.

## **Materials & Methods**

### *Hypothesis and Research Questions*

The general hypothesis underlying our research was:

*Incorporating Systems Thinking increases the effectiveness of sustainability education.*

The specific research questions were, regarding sustainability education:

**RQ1:** Does Systems Thinking enhance the learner's practical understanding of sustainability?

**RQ2:** Does interacting with System Dynamics simulations enhance the learner's practical understanding of sustainability?

**RQ3:** Does adding *both* Systems Thinking and System Dynamics simulation enhance learning more than Systems Thinking only, simulation only, or a non-systemic treatment?

**RQ4:** Do Systems Thinking and/or simulation support the transfer of sustainability understanding from one problem to another with a similar systemic structure?

A brief account of the initial design of the study was published before the study was conducted [45]. A fuller account of the design, together with results and analysis, are all documented in the following sections.

#### *Study Design*

The study concerned comparison of educational outcomes, therefore the design was drawn from established practices in Social Sciences research [46]. The investigation was an experimental study using a two-by-two factorial design. The two factors, Systems Thinking and simulation, each had two levels, present or absent. To answer the research questions, the study aimed to discover the main effects, i.e. the effect of each factor on the learning outcome, and the interaction effect, or the combined effect of both factors.

The study was conducted in the summer of 2020. Participants were randomly assigned to one of four groups: a control group, a Systems Thinking (ST) group, a Simulation (Sim) group, and a Systems Thinking and Simulation (ST + Sim) group (see Table 1). They were then given access to an online learning tool. The control group saw only standard, non-systemic content. The other groups saw additional content according to their group, either a Systems Thinking section, a simulation section, or both. All groups took the same two quizzes, and the performance of the groups in these quizzes was compared using statistical methods.

Table 1 A two-by-two factorial design resulting in four experimental groups

Factors	No Systems Thinking	Systems Thinking
No Simulation	Control group	ST group
Simulation	Sim group	ST + Sim group

#### *Teaching method*

The learning tool was originally planned to be used in a small group classroom context, with the researcher delivering an overview to the whole group before each participant would then engage with the learning tool individually. The researcher would have been available in person to answer questions about how to use or navigate the tool or to resolve any technical issues that might have arisen. However, due to Covid-19 restrictions, the training was re-designed as a single online unsupervised individual session. Support was available from the researcher via email.

The learning session lasted between approximately 50 minutes for the control group and 100 minutes for the full treatment (ST + Sim) group.

#### *The Sustainability Learning Tool*

An open access version of the learning tool is available here: <https://exchange.iseesystems.com/public/carolineb/sustainability-learning-tool><sup>3</sup>. It is not a simulation game or a flight simulator, in that

<sup>3</sup> It differs from the original version used for research in that it does not require a login and does not collect login or simulation data. Survey answers, although recorded, will no longer be checked, and may be deleted after

learners are not asked to take the role of an actor in the scenarios. The Systems Thinking and simulation elements in the learning tool offer the 'big picture' of the systems underlying two sustainability problems and offer insights into their essential structure and dynamics. The learning tool also explores sustainable solutions to the problems. The emphasis is thus on systemic understanding and policy making.

The learning tool consists of two main sections, one for deer herd management and one for fisheries, as shown in Figure 1. The deer section contains additional sections for Systems Thinking and simulation.

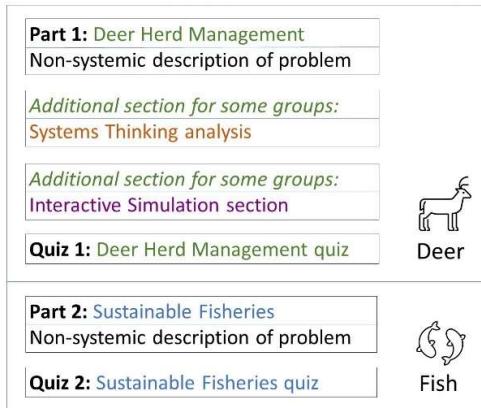


Figure 1 Sections of the ESD learning tool

The learning tool features are summarised in Table 2. The design elements are outlined briefly in the following sections, except for SBLE design principles, which were discussed in the introduction.

Table 2 Elements of the Learning Tool

Design element	Description
Overall features	<ul style="list-style-type: none"> <li>Online, single session, lasting 50 to 100 minutes</li> <li>Standard non-systemic introduction to the main topics using text, images, graphs and embedded short video clips and domain-specific terminology</li> <li>Embedded quizzes at the end of the two main topics to measure learning, and surveys to gather subjective feedback on features</li> </ul>
Case Studies	<ul style="list-style-type: none"> <li>Deer Herd management (illustrated with the story of Kaibab deer)</li> <li>Sustainable fisheries (illustrated with the story of Grand Banks cod fishery collapse)</li> </ul>
Systems Archetype	<ul style="list-style-type: none"> <li>Limits to Growth (Overshoot and Collapse).</li> <li>Applies to both case studies</li> </ul>
System Dynamics model	<ul style="list-style-type: none"> <li>The model underlying the simulation exercises was a deer model adapted from Breierova [47] and created in Stella Architect</li> </ul>
SBLE design principles	<ul style="list-style-type: none"> <li>Case-based learning</li> <li>Tasks, exercises and feedback guide use of simulations</li> <li>Encourage reflection and promote cognitive dissonance (reconstruct beliefs)</li> <li>Model/simulations increase in complexity</li> <li>Small model</li> </ul>

30 days. The pre-survey, which included a consent form, has been removed. Learners are no longer randomly allocated to treatment groups, but instead can choose which version of the tool they wish to see.

<b>Sustainability Principles or topics</b>	<ul style="list-style-type: none"> <li>• Learner can change some model variables and see graphical results instantly</li> </ul> <p>Sustainability skills/knowledge:</p> <ul style="list-style-type: none"> <li>• General definition of sustainability</li> <li>• Sustainable use of renewable resources: limit growth, respect carrying capacity and monitor the system</li> <li>• Herman Daly principle: Renewable resources must be used no faster than the rate at which they regenerate</li> <li>• Perform growth calculations and interpret graphs</li> <li>• Define sustainability in context</li> <li>• Understand limits and capacity (includes carrying capacity, maximum sustainable herd size, maximum sustainable yield in fisheries, overgrazing and overfishing)</li> <li>• Dynamic reasoning (including stock and flow reasoning)</li> <li>• Analyse a complex system (structural causes of dynamic behaviour)</li> <li>• Make judgments about sustainability (whether a system is sustainable or not)</li> <li>• Strategies for sustainability</li> </ul>
<b>Systems Thinking Principles or topics</b>	<p>Systems Thinking skills/knowledge:</p> <ul style="list-style-type: none"> <li>• Define systems and Systems Thinking</li> <li>• Feedback loops</li> <li>• Causal Loop diagrams</li> <li>• Behaviour over Time graphs</li> <li>• Structure determines behaviour</li> <li>• Stock and Flow diagrams</li> <li>• Identify common system patterns (archetypes)</li> <li>• Identify leverage points (places to intervene in a system)</li> <li>• Understand system equilibrium (a dynamic and sustainable state)</li> </ul>
<b>Simulation Exercises</b>	<ul style="list-style-type: none"> <li>• Simulate deer herd growth in first four years (exponential increase)</li> <li>• Simulate deer herd growth in first ten years (exponential increase and then decline)</li> <li>• Simulate deer herd growth, this time with vegetation added to the graph (vegetation decline explains decline in deer population)</li> <li>• Simulate to find the estimated vegetation level after one year, given vegetation growth and simultaneous consumption by deer (interacting stock levels are hard to calculate without simulation)</li> <li>• Try lowering initial deer population to avert collapse (this only delays it)</li> <li>• Try increasing initial vegetation level to avert collapse (this only delays it)</li> <li>• Try changing deer birth and death rates to obtain a stable population (birth and death rates must be equal)</li> <li>• Try to make the deer herd sustainable (stabilise deer population AND ensure it does not exceed the carrying capacity)</li> </ul>

#### Standard non-systemic introductory pages

Introductory pages consist of text, images, graphs and short embedded video clips. The general concept of sustainability is first explained and explored, then each sustainability theme is described using standard domain-specific terminology. See Figure 2 and Figure 3 for sample pages, one for each case study.

Intro  
3 of 9

## Sustainability Challenge 1: Deer Herd Management

Imagine you are managing a deer herd. The story below describes what you want to avoid - a herd that grew unsustainably and then collapsed.

### The Story of the Kaibab Deer

In 1905 the deer population on the Kaibab Plateau in Arizona, USA, was about 4,000. In 1906, a new policy led to the elimination of both deer predators\* and hunting. As a result, the deer population exploded until it reached about 100,000 in 1924.

But because the **carrying capacity** of the land was only about 30,000, the deer began to suffer and starve in large numbers. The population collapsed, and the carrying capacity of the land was greatly reduced afterwards, damaged by overgrazing.

\* According to the records, 781 mountain lions, 30 grey wolves, 4,889 coyotes and 354 bobcats were killed.

Graph credit: <https://www.chegg.com/homework-help/questions-and-answers/for-following-graph-shows-number-kaibab-deer-isolated-kaibab-plateau-northern-arizona-declared-c0395273>

Figure 2 Sample screenshot from the introductory section (non-systemic description of the deer herd management problem)

**Fisheries**  
1 of 7

## Sustainability Challenge 2: Fisheries Management

**How do you manage a fishery sustainably?**

**What is a fishery?**



A fishery is either a particular region or a particular species of fish, and usually both. Examples are the salmon fishery of Alaska or the tuna fishery of the Eastern Pacific.

Here's an example of an unsustainably managed fishery:

**A famous example of fishery collapse**

For one thousand years the Grand Banks cod fishery off the coast of Newfoundland, Canada, had been one of the richest fishing grounds in the world. Serious overfishing began in the late 1950s, which, together with poor management, led in 1992 to the total collapse of the fishery. A moratorium (a temporary ban) was declared, devastating the local community and industry. Damage done to the coastal ecosystem proved irreversible, however, and the cod fishery remains closed to this day.

This section:

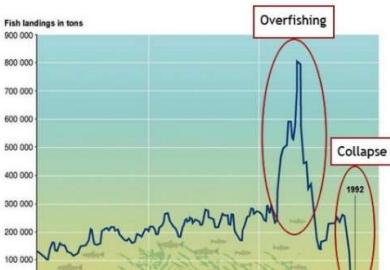


**Overfishing**

**Collapse**

1992

**Fish landings in tons**



Graph source: Millennium Ecosystem Assessment Synthesis Report (2005), Chapter 3, p.58



Figure 3 Sample screenshot of Part 2 (Non-systemic description of sustainable fisheries management)

## Embedded Quizzes and Surveys

Quiz 1 (deer management) and quiz 2 (fisheries) were tests of ability, used to measure sustainability understanding of the deer and fisheries sustainability problems. These quizzes were embedded in the tool as shown in Figure 4. All participants took these quizzes, and the pre-survey, containing the consent form. Some groups saw short post-ST and simulation feedback surveys, which were also embedded in the tool. Table 3 summarises the quizzes and surveys seen by each group.

Figure 4 Quiz 2 embedded in the learning tool

Table 3 Quizzes and surveys seen by each group

Group	About this Study (Pre-Survey)	Quiz 1 (Deer Quiz)	Post Deer Quiz: ST Survey	Post Deer Quiz: Sim Survey	Post Deer Quiz: ST + Sim Survey	Quiz 2 (Fisheries Quiz)
Control Group	✓	✓				✓
ST Group	✓	✓	✓			✓
Sim Group	✓	✓		✓		✓
ST + Sim Group	✓	✓			✓	✓

All quizzes and surveys were refined by pilot testing. They are openly available along with the study data in the Zenodo dataset (URL: <https://zenodo.org/record/5569508>).

#### Case Studies, Systems Archetype and System Dynamics model

The learning tool supports a teaching approach that combines sustainability topics with case studies, Systems Thinking and simulation. Exploring two specific problems increases understanding of sustainability in context. The two problems, deer herd management and sustainable fisheries, are both examples of renewable resource management. Each problem is illustrated with a historic case of over-exploitation of the renewable resource, leading to overshoot and collapse.

The catastrophic unsustainable growth of the Kaibab deer herd in the US in the 1920s has been the subject of analysis by System Dynamicists including Donella Meadows<sup>4</sup> and Andrew Ford [16] (p. 267). If natural predators are removed, deer will go on breeding until they overgraze and risk exceeding the carrying capacity of their environment.

The collapse of the Grand Banks cod fishery in 1992 is a famous example of disastrous unsustainable fishing practices [48]. Once one of the richest fishing grounds in the world, in 1992 the fishery collapsed completely, devastating the local community and economy. The collapse was caused by serious overfishing, which began in the late 1950s, together

<sup>4</sup> Donella Meadows lecture entitled 'System Dynamics Model: Kaibab Deer Population' is available online. Viewed 9 November 2021. <<https://www.youtube.com/watch?v=2rUXm5b-gZM>>

with poor management. Damage done to the coastal ecosystem proved irreversible and the cod fishery remains closed.

The Limits to Growth archetype, also known as Overshoot and Collapse [8] (p. 123), describes the behaviour of both these case studies well. The generic structure underlying this archetype consists of two stocks. The first stock grows exponentially while depending on a second stock, which is a renewable resource. Here, a fast-growing deer herd is eating ever more vegetation, and a growing fishing industry is exploiting fish stocks more and more heavily. This systemic structure will tend to cause the following behaviour. The first stock grows so rapidly that it overshoots, depleting the resource more rapidly than it can renew itself, leading to the collapse of the resource, and then the stock that depends on it. The deer herd overgrazes, causing collapse of the vegetation supply, and then the herd. The fishing industry overfishes, so that the fish population cannot reproduce itself, destroying the industry.

The remedy for this problematic dynamic is that the exponential growth of the first stock must be checked, so that the resource on which it depends will not be consumed faster than it can regenerate. If limits (e.g., carrying capacity or maximum sustainable yield) are respected, then the system can become sustainable, meaning that the second stock, the renewable resource, remains available to the first stock indefinitely because it is not overexploited and there is time for it to renew itself. A description of this strategy for sustainability was seen by all participants, including the control group.

The System Dynamics deer herd model used in the learning tool is slightly adapted from that documented by Breierova from the MIT System Dynamics in Education Project [47]. It is available in the Zenodo dataset published for this study. In her article, Breierova explains the usefulness of generic structures in helping transfer knowledge among different systems.

#### Systems Thinking principles

The essential Systems Thinking concepts, tools and techniques listed in Table 2 were chosen from the literature [49-51] as suitable tools for analysis of the two sustainability problems under consideration. These concepts, tools and techniques are explained in general terms in the first pages of the Systems Thinking section, and then they are used to analyse the deer herd population dynamics. Two sample pages from the Systems Thinking section are shown in Figure 5 and Figure 6. The section takes about 30 minutes to work through.

#### Simulation

Simulations of the deer herd population model provided the basis for a series of six exercises and two tasks, listed in Table 2, which explore the dynamics that lead to overshoot and collapse, and how those dynamics can be changed so that the herd size can become sustainable. The exercises explore in stages the interplay between the two stocks, deer and vegetation, and the key role of deer birth and death rates and vegetation regeneration and consumption rates. The exercises increase in complexity as the first stock and then the second stock is added, then the interaction between the two stocks is considered, then learners are given control of key variables so that they can explore their effects on the dynamics of the deer herd.

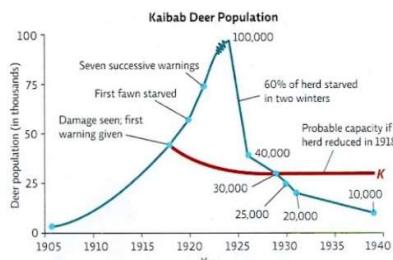
Systems  
Thinking  
10 of 20

## Sustainability Challenge 1: Deer Herd Management



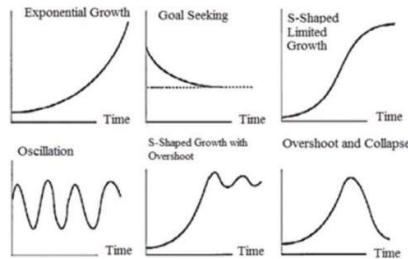
### Tool 3: Behaviour Over Time Graphs

We've already seen an example of a Behaviour over Time graph for deer, with real data:



Can we learn anything from the shape of the graph?

Here are a few common patterns of system behaviour. These same patterns show up in many, very different systems. **Which pattern (or patterns) do you think best describe the Kaibab deer population behaviour over time?**



Source: John Sterman, *Business Dynamics: Systems Thinking and Modeling for a Complex World* (2000), Chapter 4, p. 108

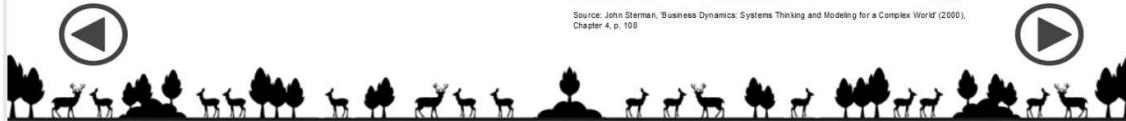


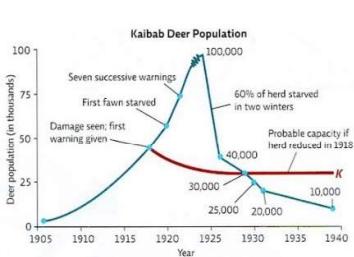
Figure 5 Screenshot of a page from the Systems Thinking section (first page of analysis of behaviour over time graph for Kaibab deer)

Systems  
Thinking  
11 of 20

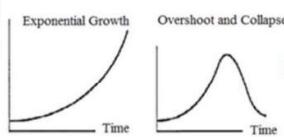
## Sustainability Challenge 1: Deer Herd Management



### Tool 3: Behaviour Over Time Graphs



This pattern seems to describe the first part of the Kaibab graph - from 1905 to 1922



This pattern is similar to the Kaibab graph as a whole - from 1905 to 1939



?

The behaviour of the Kaibab deer population starts with Exponential Growth and then turns into Overshoot and Collapse - why is that?



Figure 6 Screenshot of a page from the Systems Thinking section (second page of analysis of behaviour over time graph for Kaibab deer)

The simulation section is presented using text and embedded simulations, with a pop-up comment, hint or explanation available for each exercise to provide feedback to the learner. The section takes about 20 minutes to work through.

The aim of this sequence of challenges is to demonstrate that a sustainable deer population can result if the birth and death rates are balanced and the population remains within the carrying capacity of the available land, so that the herd size will remain stable and will consume

no more than the regenerated vegetation. Alternative strategies that might seem attractive, such as starting with a lower population or increasing the amount of vegetation (or size of the park), are shown to be ineffective, since the powerful exponential deer population growth dynamic dominates and will reach the limits of the park, albeit a little later. This exponential growth behaviour is seen to persist as long as the birth rate is greater than the death rate. This learning process is designed to encourage reflection, confronting erroneous assumptions and reconstructing beliefs.

A sample simulation exercise is shown in Figure 7.

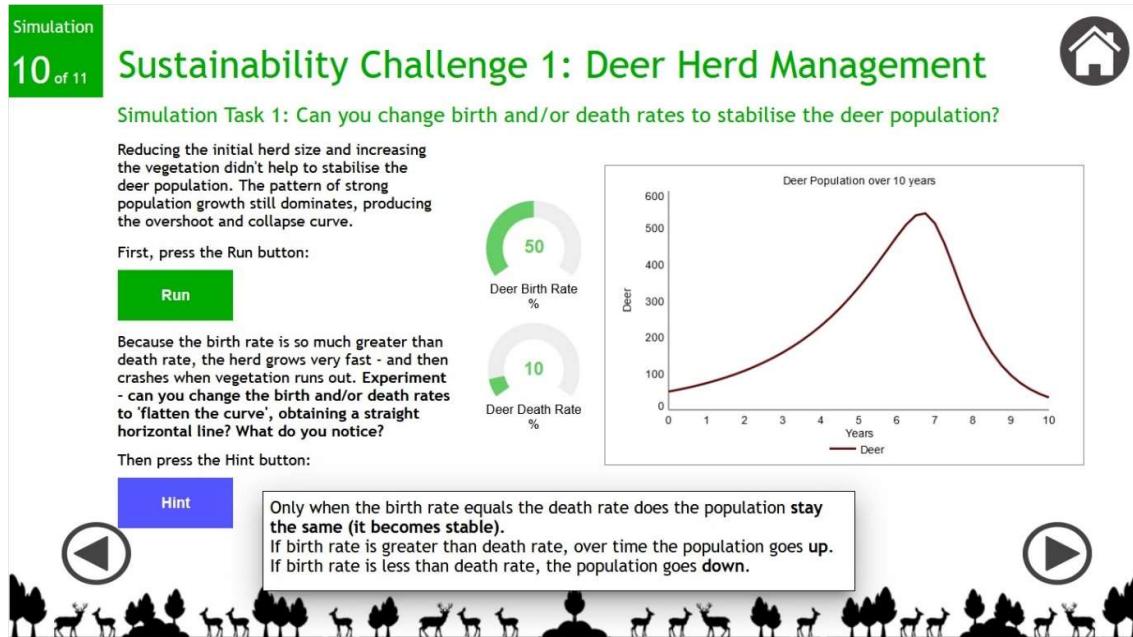


Figure 7 Sample screenshot of the System Dynamics simulations section for exploration of deer herd management problem

#### Sustainability principles

The sustainability principles and topics listed in Table 2 and tested in the quizzes were selected from the general literature on sustainability [52,53], renewable resource management [16], and a systems view of sustainability [54] (p. 214). They were chosen as necessary skills for analysing the two cases under consideration, guided by Harris [48] for the Grand Banks fishery collapse and Meadows' analysis of the Kaibab deer dynamics in her lectures, already cited. This list forms the framework for operationalising sustainability understanding using quiz 1 and quiz 2, making use of marking schemes to obtain quantitative percentage scores.

Note that these topics are limited to the cognitive aspect of sustainability understanding, not the affective, behavioural or other aspects [12].

#### Platform

Figure 8 shows the architecture of the learning tool. A gateway web page was used to allocate users to groups randomly and to provide a link to a Stella Architect interface with authentication and data collection enabled. The group id passed to the Stella interface determined

conditional pathways according to group. The Stella interface was published to the ISEE Exchange. Quizzes and surveys were embedded in the learning tool using SurveyMonkey surveys, and these employed custom variables to allow user identification taken from Stella logins. Desktop or laptops were recommended, not mobile devices, because the simulations required a reasonably large screen.

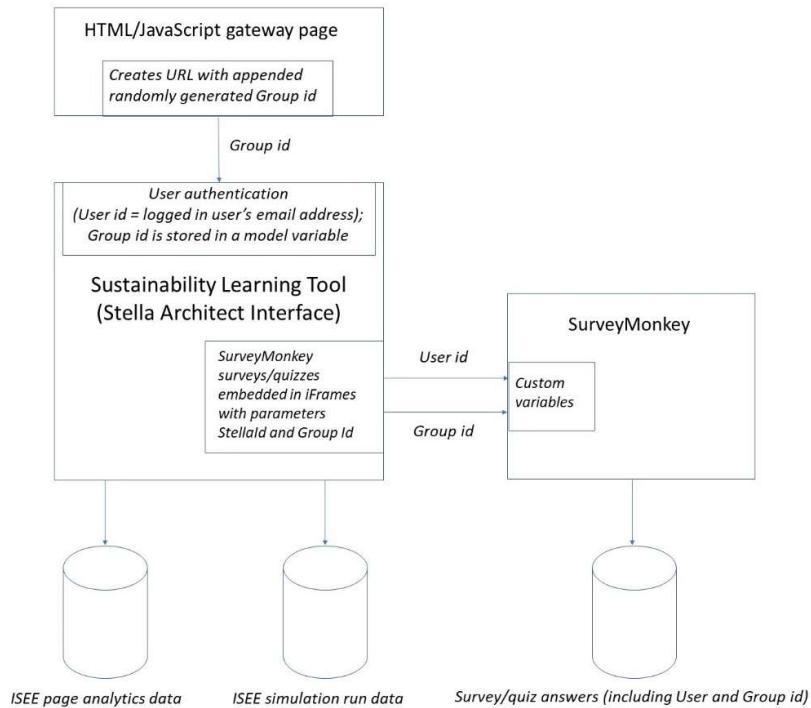


Figure 8 Learning tool architecture including data collected

#### Conditional pathways for treatment groups

The learning tool was divided into two sections, facilitating two experiments (see Figure 9). Experiment 1 was concerned with the effect of Systems Thinking and/or simulation on sustainability learning outcomes, and was designed to answer RQs 1, 2 and 3. Experiment 2 was concerned with the transfer of sustainability understanding from the deer problem to the fisheries problem and was designed to answer RQ4. Quiz 1 data were captured for experiment 1, and quiz 2 data for experiment 2.

In experiment 1, a significant increase for non-control group members in quiz 1 performance would suggest that Systems Thinking and/or simulation improved sustainability learning outcomes (RQ1, RQ2 and RQ3).

In experiment 2, a significant increase for non-control group members in quiz 2 performance would suggest that insights from Systems Thinking and/or simulation applied to the deer problem resulted in a transfer of sustainability skills to the fisheries problem (RQ4), since only a standard non-systemic description of the fisheries problem was provided.

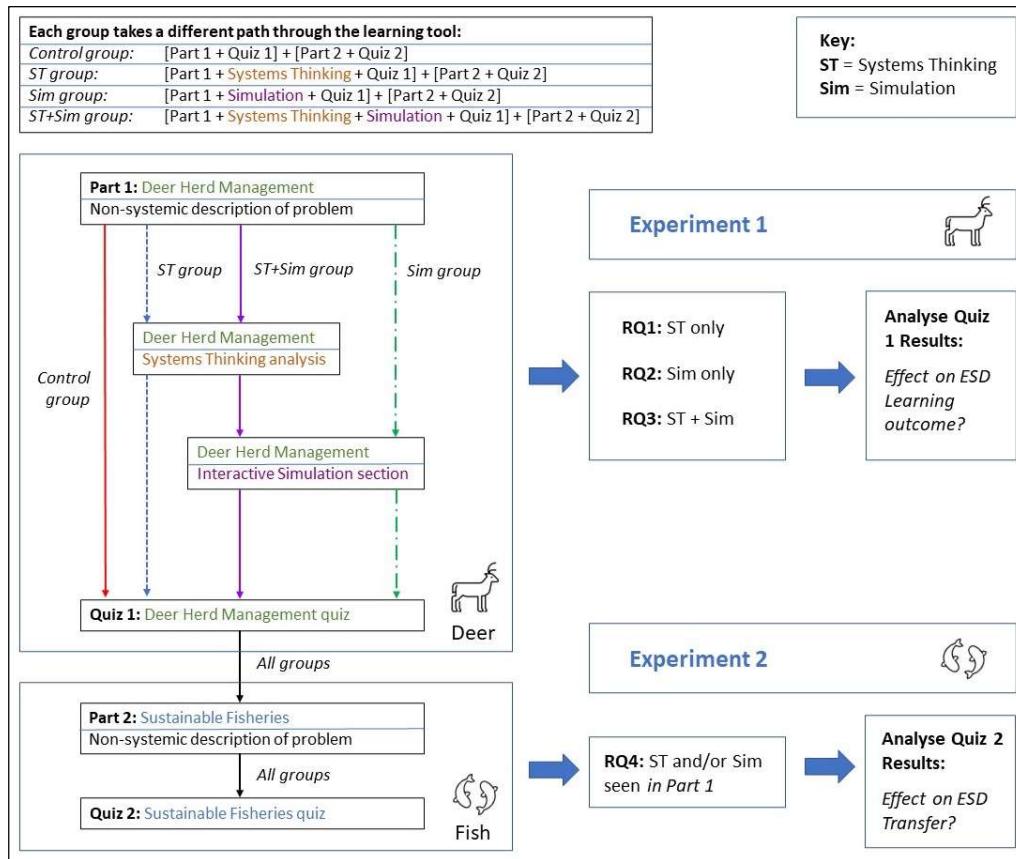


Figure 9 How the pathways through the learning tool and the experiments were designed to answer the research questions

### *Participants and sampling methods*

According to UNESCO, ESD is necessary for all ‘citizens, voters, workers, professionals, and leaders’ [55]. This is a very large population globally, so random selection was not possible because of resource and access constraints. Subjects were instead selected using non-probability sampling techniques: a combination of two forms convenience sampling with self-selection [46] (p. 113). Convenience sampling means that participants chosen were those most easily accessible. Invitations to members of the public over 18 were sent out through emails, social media or website invitations, word of mouth etc. Individuals and groups targeted included university student societies, postgraduate students, environmental organisations and political parties, friends, acquaintances and colleagues.

Those contacted were also invited to pass the invitation on to others. This is known as snowball sampling and is a form of convenience sampling. In this way the sample was extended, repeating until the required number of valid datasets was collected. Those who signed up were self-selected from this large network. A two-by-two factorial design requires a minimum of 20 participants per group [56] (p. 87), so at least 80 subjects needed to be recruited.

The Covid-19 restrictions led to a decision to deliver the learning tool for unsupervised online use. This meant that people could participate from anywhere in the world.

Randomisation was by random assignment. Whilst it is a valid method for cancelling out the effects of extraneous variables, random assignment reduces generalisability across populations when compared to random selection.

#### *Data collection, validation and anonymisation*

The following data were collected from participants:

- In the pre-survey, basic information such as age, gender, degree subject and/or occupation, and prior knowledge of sustainability.
- Quiz 1 and quiz 2 answers comprised a mix of quantitative and qualitative data, for example, numeric answers to questions about population growth, and textual answers to questions about the meaning of sustainability in context. Each quiz question was scored numerically and included in the overall percentage results. Some qualitative answers were also analysed separately.
- The short surveys, appropriate for each treatment group, collected subjective feedback about the usefulness of the simulation and Systems Thinking sections. At the end of quiz 2, participants were also asked for overall feedback about the learning tool.
- The email address used to log in to the learning tool was captured by ISEE and used to allow identification of survey, simulation and page analytics data.
  - Simulation data were used to verify that users had interacted with the simulation exercises.
  - Learning tool page analytics were used to judge whether participants engaged adequately with the learning tool.

Once the data were collected, datasets were validated. Validation rules used to define acceptable engagement with sections and delay in recording quiz answers are detailed in the codebooks available in the Zenodo dataset. Datasets were also checked for completeness, compared with surveys and quizzes for each group member summarised in Table 3. Participants were asked to complete promptly any feedback surveys that failed to record. Some cross-checking was necessary between SurveyMonkey data and ISEE data where items of data failed to record, for unknown reasons.

Login email ids in the data were replaced with anonymised participant ids to avoid rater bias [57] (p. 209). In this study, since the researcher knew some participants personally, and had access to demographic data collected in the pre-survey, there was a risk of rater bias. The researcher may also have been influenced, consciously or unconsciously, by knowing the participant's treatment group. To reduce these risks, quizzes were marked 'blind', i.e. the researcher did not know the participant's identity nor which group they were allocated to.

Quiz scores were calculated, and background variable values recorded, using predetermined marking schemes and scales. Quiz answers, marking schemes and code books for recording results are all included in the published Zenodo dataset.

#### *Data Analysis*

A Factorial ANOVA is an appropriate overall test for exploring the causal relationship between the two categorical independent variables and one quantitative dependent variable [58]. It detects whether any group differs significantly from the others. Factorial ANOVA differs from the standard ANOVA test, in which there is only one independent

variable. Certain assumptions about the distribution of the data must be met in order to conduct either form of ANOVA test. Kruskal-Wallis is a suitable alternative non-parametric overall test that can be used instead of the standard ANOVA if these assumptions are not met. If the overall test finds that there is a difference between the groups, individual post-hoc tests can be conducted. In this study the unpaired two-samples Wilcoxon tests (a non-parametric post-hoc test that can be performed after Kruskal-Wallis) and an independent two-sample t-test were conducted.

The statistical programming language R was used to create descriptive statistics such as graphs and summary statistics, to check assumptions for parametric tests, to carry out all the inferential statistics tests and to calculate effect sizes [59]. The R scripts necessary to reproduce all the results in detail are openly available in the Zenodo dataset (URL: <https://zenodo.org/record/5569508>).

Each significance test result documented in later sections includes a p-value, but arguably this does not measure the strength of the relationship. An effect size such as Cohen's  $d$  is a useful complement [60]. Cohen provided basic guidelines for interpreting the effect size, namely 0.2 as small, 0.5 as medium, and 0.8 as large [61]. However, he advised that his benchmarks were recommended for use only when no better basis is available. In education research, the average effect size is  $d = 0.4$ , with 0.2, 0.4 and 0.6 considered small, medium and large effects [62].

Randomisation in the study design aims to generate comparable groups to eliminate the effect of extraneous variables, but it is always possible that unidentified confounding variables exist, confounding can be introduced by inappropriate adjustments, and the effects of confounders may not be entirely removed [63]. The approach taken to analysing extraneous variables in this study was, where known, to check their actual distribution across groups, to see if this was even, and/or to examine their effect on scores, using stratification. Data were not formally adjusted to compensate for their effects, if found; instead, sometimes adjustments were estimated, but more generally, limitations to findings and recommendations for further experimental studies were reported.

## Results

### *Profile of participants*

Of 227 people who signed up to participate, 80 did not follow up, and 8 started but withdrew. There were 33 incomplete or invalid datasets. Some participants experienced technical issues or otherwise needed support to complete the learning experiment. After data validation there were 106 complete datasets, one dataset per participant.

The majority of participants (58.5%) were female, and 41.5% were male. The average age was around 50 years. The great majority (85%) resided in the Ireland or the UK (60% and 25% respectively), and 15% elsewhere. Most were graduates or postgraduates (77%), the average a little below Master's degree level. The majority (62%) of participants had little or no prior knowledge about sustainability. The vast majority (87%) of participants had no prior knowledge of Systems Thinking or System Dynamics. For nearly two-thirds (64%) of participants, their occupation and/or education had no relevance or little relevance to sustainability or systems thinking.

### *Descriptive Statistics*

#### Experiment 1: Quiz 1 (Deer Herd Management) Scores

The simulation group, highlighted in Table 4, performed best. All treatment groups performed better than the control group. A boxplot showing the distribution of quiz scores is shown in Figure 10. This shows an outlier in the control group.

Table 4 Deer Herd Management Quiz (Quiz 1) Scores by Group

Group	Control	ST	Sim	ST + Sim	All
<b>Total participants</b>	28	26	24	28	106
<b>Min % score</b>	40	55	48	46	40
<b>Max % score</b>	92	97	96	97	97
<b>Mean % score</b>	70.8	75.8	78.4	72.9	74.3
<b>Median % score</b>	73.5	76.5	84	77	77
<b>Standard deviation</b>	11.0	11.3	14.1	11.8	12.2
<b>Outlier score(s)</b>	40	-	-	-	-

The mean score of the full treatment group (ST + Simulation) was lower than that of the Systems Thinking group, and the Simulation group, which was unexpected. The question arose, why was the mean score obtained when both factors were combined not at least as high as that obtained with either of the factors alone?

Scores for individual quiz questions were compared to find out which groups performed best on specific sustainability topics. The ST group outperformed other groups in questions about maximum capacity. The Sim group outperformed other groups in questions about the definition of sustainability, identifying sustainable graph patterns, calculating multiple interacting stock levels, identifying the point where limits were reached, and choosing policies for sustainability.

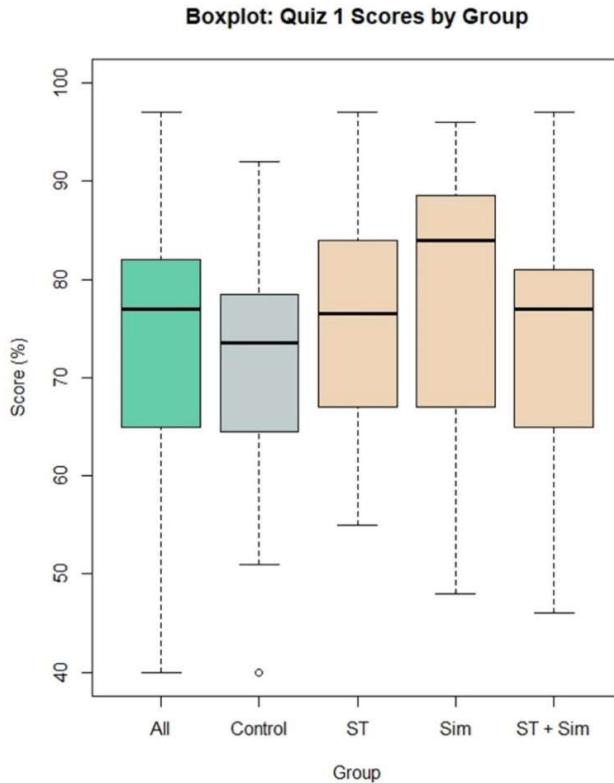


Figure 10 Boxplot of Quiz 1 Scores by Group

#### Possible reasons for poor ST + Simulation group performance

The ST + Sim group got lower than expected results. The distribution of known background variables between groups was explored in case these were confounding, and this group was found to differ from the others in three ways. There were far more over 65s, they had far less prior sustainability knowledge, and there were far more delays in both quizzes due to technical issues or interruptions. However, after closer analysis using stratification, higher age and more delays were found not to be associated with lower quiz scores. A lower average prior sustainability knowledge score did affect score a little: estimating the effect on group mean score of increasing the average level of prior knowledge to that of other groups suggests an increase of 1.2%, not enough to create a significant result for that group. However, further studies could use techniques such as restriction or matching in the study design to eliminate any possible effect. A design that simply excluded people with high prior sustainability knowledge could be sufficient.

A much more likely explanation for the poor performance of this group, the full treatment group, was found in the significant negative interaction effects uncovered by Factorial ANOVA testing and described in the Inferential Statistics section.

#### Experiment 2: Quiz 2 (Sustainable Fisheries Management) Scores

Again, the simulation group, highlighted in Table 5, performed best. Other treatment groups performed worse than the control group, when comparing mean, median and mode of group scores. A boxplot showing the distribution of quiz scores is shown in Figure 11. There are two

outliers in the control group, one of which (the lowest score) belonged to the same participant as the outlier in Figure 10. One outlier was less extreme, and was only an outlier for the control group, not the participants as a whole. No obvious errors or unusual circumstances gave rise to this outlier.

Table 5 Sustainable Fisheries Quiz (Quiz 2) Scores by Group

Group	Control	ST	Simulation	ST + Simulation	All
<b>Total participants</b>	28	26	24	28	106
<b>Min % score</b>	38	55	55	54	38
<b>Max % score</b>	92	97	95	89	97
<b>Mean % score</b>	78.3	77.5	82.0	73.7	77.7
<b>Median % score</b>	80.5	78	83.5	73.5	79.5
<b>Standard deviation</b>	11.1	11.2	9.9	8.9	10.6
<b>Outlier score(s)</b>	38, 59	-	-	-	38

Boxplot: Quiz 2 Scores by Group

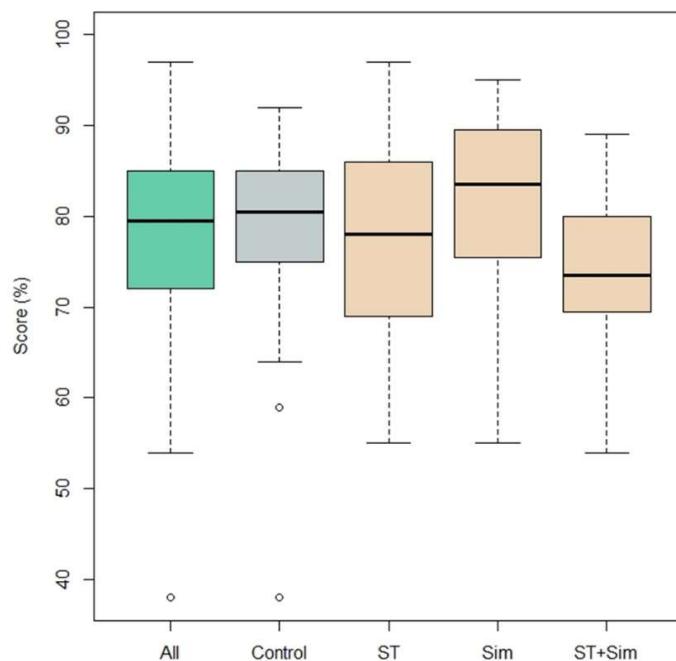


Figure 11 Boxplot of Quiz 2 Scores by Group

Scores for individual questions were compared by group to find out which groups performed best on specific sustainability topics. The Sim group outperformed all other groups in the calculation of years to maximum fishery capacity, in understanding maximum sustainable yield, and in identifying sustainable graph patterns. It performed a little better than other groups in single stock exponential growth and maximum capacity calculations, and in defining sustainability in the context of fisheries.

### *Inferential Statistics*

Table 6 summarises the process followed when conducting the inferential statistical tests on quiz 1 and quiz 2 data. The main findings are highlighted. All R scripts necessary to reproduce the results outlined here, including the analysis of possible confounding variables, are available in the Zenodo published dataset. The following paragraphs provide explanatory notes.

From the literature, there is a clear expectation that Systems Thinking and/or simulation will increase understanding of sustainability problems. Therefore, alternative hypotheses tested asserted that scores for treatment groups would be greater than those of the control group, leading to right-tailed (one-tailed) significance tests. The null hypotheses were that there were no differences between the groups.

Where a parametric test was conducted, the appropriate assumptions for the test were first checked. The assumptions for ANOVA tests are the independence of observations, the homogeneity of variances and the normality of residuals. The first condition is satisfied since participants in this study were randomly allocated to treatment groups. Levene's test for homogeneity of variance and the Shapiro-Wilk test for normality of residuals were both carried out using R on the appropriate datasets to check the other two assumptions. The assumptions for the two-sample independent t-test are similar: independence of the observations, an approximately normal distribution for each group, and homogeneity of variances. If the assumptions were not met, a suitable non-parametric test was used instead.

Parametric tests do not work well when there are outliers [46] (p. 592). The outlier score in quiz 1 was removed, and the more extreme outlier in quiz 2 removed, before ANOVA testing. Finally, two quiz 2 datasets were removed prior to analysis, as page analytics logs revealed that these participants did not engage with the fisheries section of the learning tool. They both spent no more than two minutes on the fisheries section, whereas the minimum acceptable time was 5.5 minutes, and recommended time was 15 minutes.

The Factorial ANOVA tests revealed that both in quiz 1 and quiz 2 the presence of both factors (Systems Thinking and simulation) created a negative, or 'antagonistic' interaction effect. Interaction plots are shown in Figure 12 and Figure 13. This means that adding a second treatment reduced the quiz scores. The interaction effect partly cancelled out the main effects of each factor alone. This refutes RQ3.

### *Effect Size*

The best performing group was the Sim group with Cohen's d effect size calculated at 0.6. This is a large effect in an educational context. ST improved learning outcomes but had a weaker effect (Cohen's d 0.4, a medium effect). ST + Sim had a still weaker effect (0.1, a very small effect).

For quiz 2, the Sim group was the only group that performed better than the control group, so the effect size for other groups was not calculated. Cohen's d was calculated at 0.4. This is a medium effect in an educational context.

Table 7 provides a formal summary of the results of the inferential tests and effect sizes and provides answers to the research questions and main hypothesis.

Table 6 Overview of inferential testing process

Quiz data	Steps	Purpose	Type	Test	Result	Interpretation	Decision
Quiz 1	1	Overall test on all 4 groups	Parametric	Factorial ANOVA	A significant result only for the interaction of ST and Sim (p value .045), not for either of the factors alone (the main effects).	An interaction plot shows an 'antagonistic' interaction effect, very strong because the lines are nearly perpendicular. Main effects are therefore uninterpretable [56].	Remove data for ST + Sim group. Use a one-way ANOVA (group as independent variable)
	2	Overall test on 3 groups (control, ST and Sim)	Non-parametric (assumptions not met for ANOVA)	Kruskal-Wallis test	p-value .067, significant at the 90% confidence level.	It is likely that at least one group differs from the others.	Proceed with post-hoc tests to compare ST & control, and Sim & control (2 comparisons)
	3	Post-hoc tests on the above three groups	Non-parametric (assumptions not met in step 2)	Pairwise Wilcoxon rank sum tests (one-tailed)	One significant result, an unadjusted p-value 0.009 for the Sim group. The adjusted p-value of .018 (Bonferroni adjustment for 2 comparisons) is significant at the 95% confidence level.	The Sim group scored significantly better than the control in quiz 1.	- End testing -
Quiz 2	1	Overall test on all four groups	Parametric	Factorial ANOVA	A significant result for the interaction of ST and Sim (p-value .052) at the 90% confidence level.	An interaction plot shows an 'antagonistic' interaction effect. The lines cross and both slope downwards. The downward slope shows that ST always brings down the score.	Only Sim group performed better than control. Proceed to compare these.
	2	Compare Sim and control groups	Parametric	Two-sample independent t-test (one-tailed)	p-value .079, significant at the 90% confidence level.	Sim scores were better than the control group scores in quiz 2, but the result is weaker than for quiz 1.	- End testing -

Interaction plot for ST and Simulation (Quiz 1)

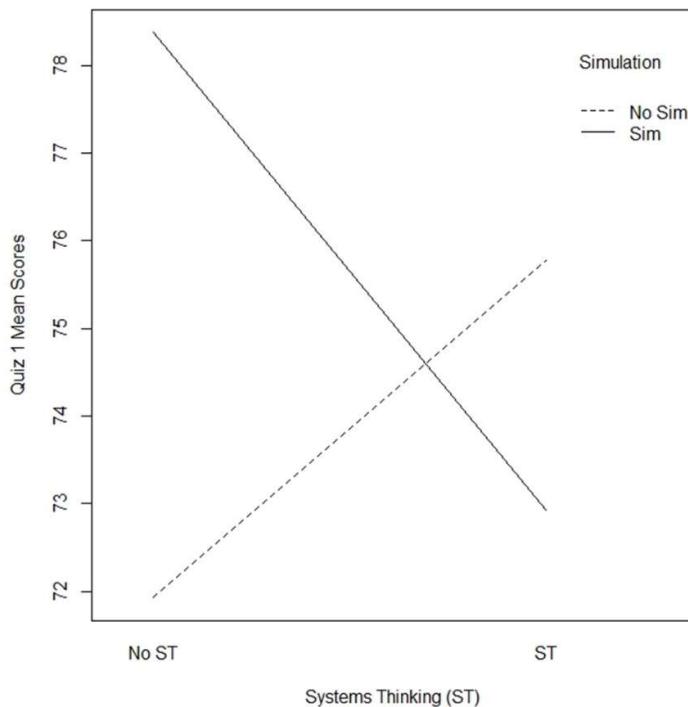


Figure 12 Interaction plot showing effect of both factors on quiz 1 score

Interaction plot for ST and Simulation (Quiz 2)

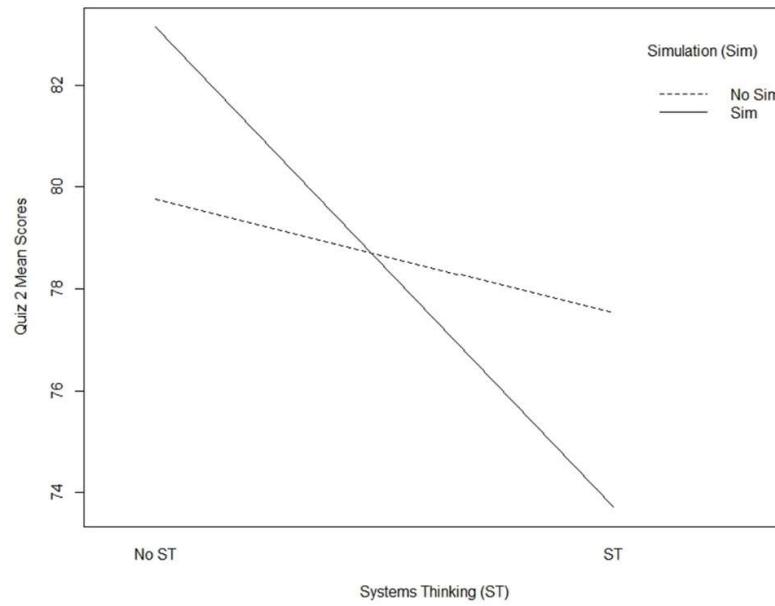


Figure 13 Interaction plot showing effect of both factors on quiz 2 score

Table 7 Summary of results of inferential testing

Research Question	Summary of Results of Inferential Statistical testing
<b>RQ1:</b>	<b>Does Systems Thinking enhance the learner's practical understanding of sustainability?</b>
<i>Result:</i> <b>No</b>	A Wilcoxon rank sum test (one-tailed, and using the Bonferroni correction to adjust p) showed that there was no significant increase in mean scores for the 26 participants in the Systems Thinking group ( $M = 75.8$ , $SD = 11.3$ ) compared to the 27 participants in the control group ( $p = .247$ ), despite participants attaining higher scores than the control group ( $M = 71.9$ , $SD = 9.3$ ). The effect size was medium in the educational context (Cohen's $d = 0.4$ ).
<b>RQ2:</b>	<b>Does interacting with System Dynamics simulations enhance the learner's practical understanding of sustainability?</b>
<i>Result:</i> <b>Yes</b>	A Wilcoxon rank sum test (one-tailed, and using the Bonferroni correction to adjust p) showed that the 24 participants in the Simulation group ( $M = 78.4$ , $SD = 14.1$ ) compared to the 27 participants in the control group ( $M = 71.9$ , $SD = 9.3$ ) demonstrated significantly better mean scores, ( $p = .018$ ) at the 95% confidence level ( $\alpha = 0.05$ ). The effect size was large in the educational context (Cohen's $d = 0.6$ ).
<b>RQ3:</b>	<b>Does adding both Systems Thinking and System Dynamics simulation enhance learning more than Systems Thinking only, simulation only, or a non-systemic treatment?</b>
<i>Result:</i> <b>No</b>	The 28 participants in the Systems Thinking and Simulation group performed slightly better ( $M = 72.9$ , $SD = 11.8$ ) than the control group ( $M = 71.9$ , $SD = 9.3$ ), but worse than the other treatment groups. The effect size was very small in the educational context (Cohen's $d = 0.1$ ). A Factorial ANOVA test found a significant interaction effect between the two factors on score ( $p = .045$ ). The interaction effect was negative, since the presence of one factor reduced the effect of the other. The interaction effect was therefore antagonistic to the main effects of the factors.
<b>RQ4:</b>	<b>Do Systems Thinking and/or simulation support the transfer of sustainability understanding from one problem to another with a similar systemic structure?</b>
<i>Result:</i> <b>Yes, only simulation</b>	The 23 participants in the Simulation group ( $M = 83.1$ , $SD = 8.3$ ) compared to the 26 participants in the control group ( $M = 79.8$ , $SD = 8.1$ ) did not demonstrate significantly better mean scores at the 95% confidence level ( $\alpha = 0.05$ ): $t(47) = 1.44$ , $p = .0787$ . However, the mean scores were significantly better at the 90% confidence level ( $\alpha = 0.10$ ). The effect size was medium in the educational context (Cohen's $d = 0.4$ ). Other treatment groups did not perform better than the control group. A Factorial ANOVA test found a significant interaction effect at the 90% confidence level between the two factors on score ( $p = .052$ ). The interaction effect was negative, since the presence of each factor reduced the effect of the other factor. Systems Thinking reduced scores.
<b>Hypothesis:</b>	<b>Incorporating Systems Thinking increases the effectiveness of sustainability education.</b>
<i>Result:</i> <b>Only simulation</b>	Only simulation was found to significantly increase sustainability quiz scores. Systems Thinking increased scores in quiz 1, but not significantly.

## *Feedback from Participants*

The quantitative and qualitative feedback data summarised in this section are available in full in the published Zenodo dataset. The main points are summarised below.

## General Feedback on the Learning Tool

Comments were generally very positive. The most frequent words used are visualised in the word cloud in Figure 14. The most frequent evaluative words used were 'interesting' and 'informative'.

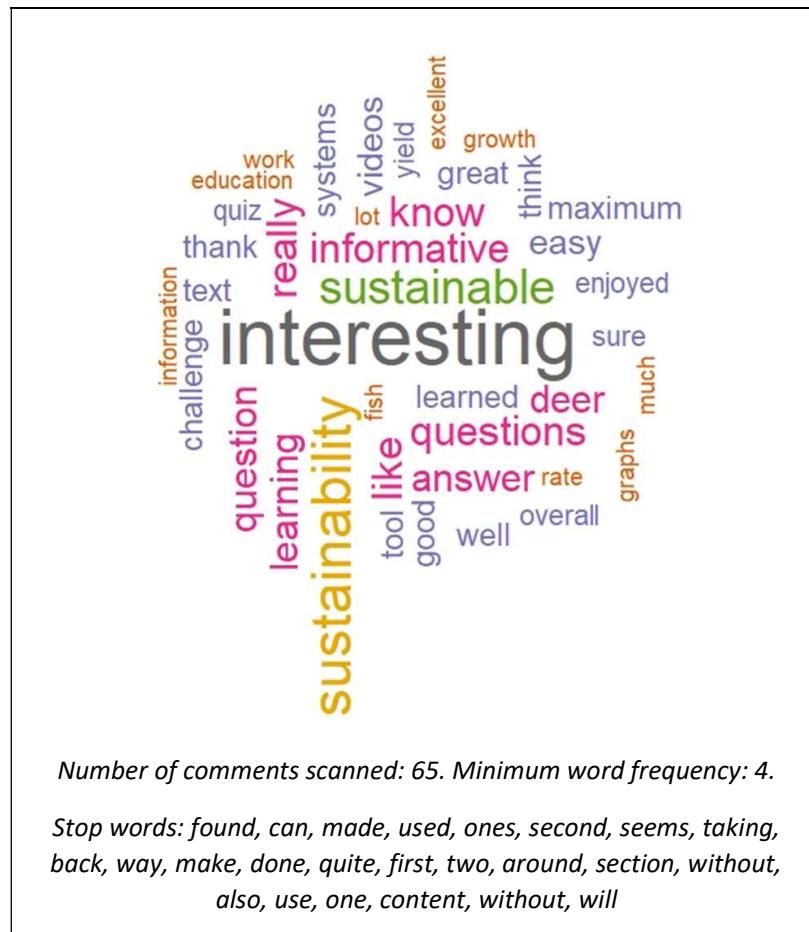


Figure 14 Word cloud of optional general comments about the learning tool

Summarising the 65 comments, some people commented favourably on the benefits of interactive learning with simulation. It helped them better understand cause and effect and consequences of policy decisions, allowed experimentation, knowledge construction, and made learning enjoyable. Case studies were found useful.

Some people remarked that they understood the first problem better than the second, because of the Systems Thinking analysis and/or simulation provided for the first problem and not the second. Perhaps they struggled to transfer their understanding from the first problem to the second, unlike this person, who saw the connection: 'The similar nature of the two problems meant that lessons learned in simulations on the first problem applied to the second problem without simulations'. Another person said that they internalised the lessons from simulation

and were able to apply those lessons to the second problem without explicit simulation material.

This comment suggests the usefulness of combining both factors: 'The challenge in education is that often times these two valuable tools (systems thinking and simulation) are parsed but their combined use here has been excellent in facilitating knowledge construction'.

Some people found the mathematical aspects of the learning material challenging, and another cautioned about the complexity of the Systems Thinking section.

#### Feedback on the usefulness of Systems Thinking

Participants with access to the Systems Thinking section were asked to rate how useful they found it, on a 5-point Likert scale (see Figure 15). About three-quarters (74.1%) of participants said it helped quite a lot or really transformed the way they saw the problem.

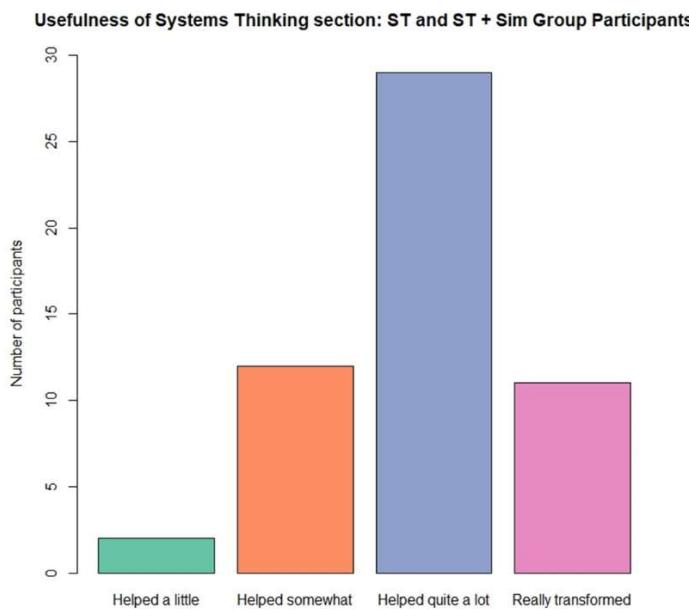


Figure 15 Bar graph showing usefulness ratings given for Systems Thinking

The most frequent words found in the 24 optional comments given about the usefulness of Systems Thinking are visualised in Figure 16. The most frequent evaluative words were 'helped', 'helpful' and 'excellent'. Participants remarked on the usefulness of graphs, diagrams, videos and highlighted words for simplifying a complex topic and making it more memorable. Systems Thinking was found useful for understanding interrelationships and how systems interlink, identifying the point at which systems become unsustainable, clarifying cause and effect, identifying patterns of behaviour and changes over time, and making decisions. A few expressed concern about remembering the complex terminology.

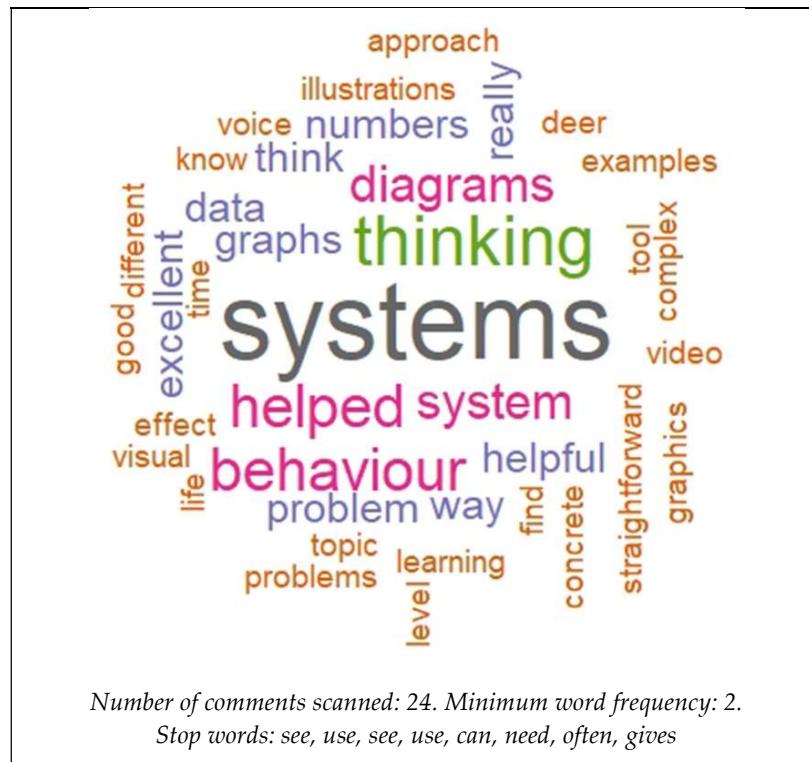


Figure 16 Word cloud of optional feedback comments on Systems Thinking

#### Feedback on the usefulness of Simulation

Participants with access to the simulation section were asked to rate how useful they found it on a 5-point Likert scale (see Figure 17). The great majority (84.6%) of participants felt that simulation helped quite a lot or really transformed the way they saw the problem, higher even than the 74.1% of participants who felt the same about Systems Thinking. When analysed by group (figures not shown), 92% of Sim group participants felt that simulation helped quite a lot or really transformed the way they saw the problem, compared with 78.6% of ST + Sim group participants.

Usefulness of Simulation section: Sim and ST + Sim Group Participants

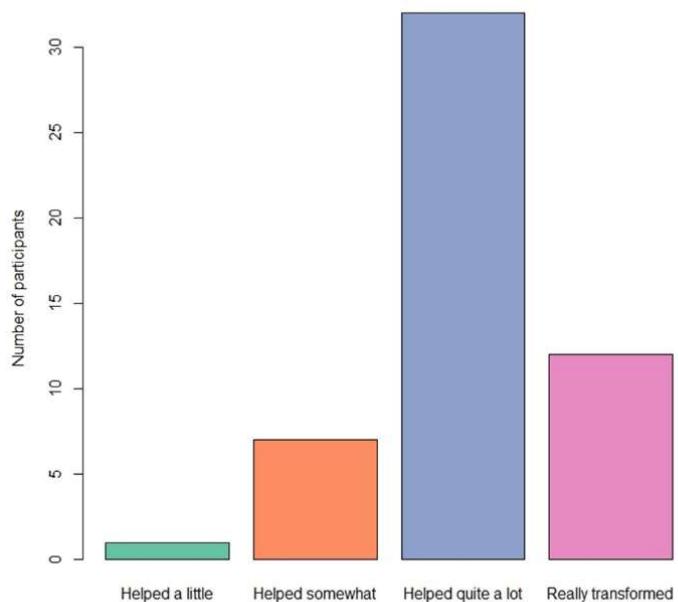


Figure 17 Bar graph showing usefulness ratings given for Simulation

The most frequent words found in the 32 optional comments about the usefulness of simulation are visualised in Figure 18. The most frequent evaluative or descriptive words used were 'useful', 'interactive', 'learning', 'remember', 'understanding' and 'thinking'. Participants found simulation useful for increasing clarity and understanding by adjusting variables, experimenting with strategies, assessing impacts and informing policy decisions, for seeing how quickly resources can be depleted, for finding sustainable limits, and for teaching responsibility. Interactivity helps learning and retention, some said, and seeing graphs change dynamically is more effective than reading text for understanding complexity and real-world problems and performing the mathematical work themselves.

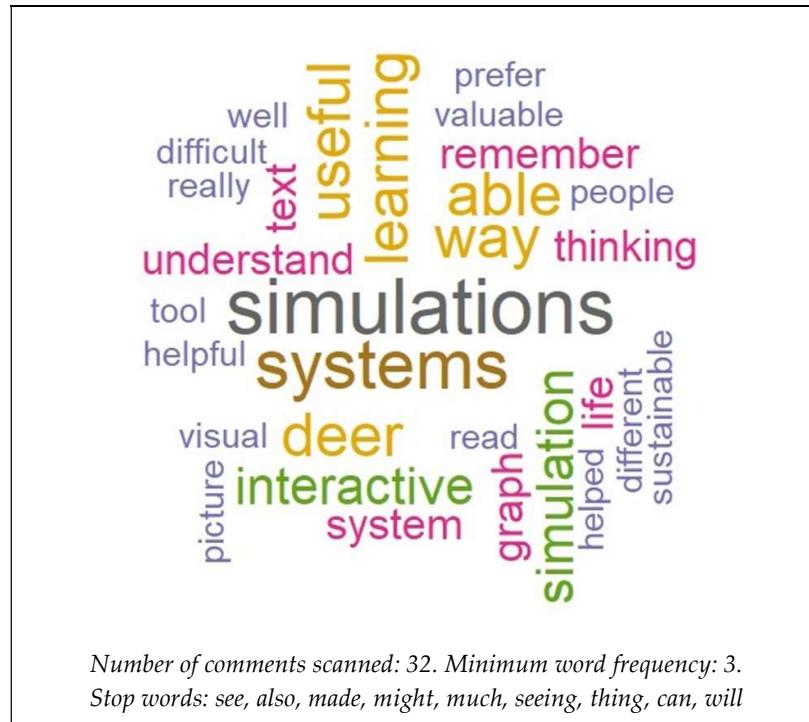


Figure 18 Word cloud of optional feedback comments on Simulation

### Discussion and Conclusions

The main findings were that System Dynamics simulation has a strong effect on understanding a sustainability problem, and a weaker but significant effect on transfer of understanding to another problem with a similar systemic structure. Systems Thinking did not make a significant difference to mean scores in either case, and the combination of Systems Thinking and simulation in the full treatment group proved negative. This could be evidence that the additional learning material, or perhaps its abstract complexity, pushed participants over a limit with respect to 'cognitive load' [64] in this experimental setting (a single learning session). It could also be evidence that quantitative simulation has a better learning outcome than more qualitative approaches. Interactive simulation provides an opportunity for learners to perform actions (operations) and build their understanding of a system through 'operational thinking' [31,65].

Feedback from participants was very positive with a large majority reporting finding both Systems Thinking and simulation useful.

In conclusion, simulation is a highly effective tool for enhancing sustainability understanding in a single short learning session, even when learning is done remotely online without supervision.

#### *Limitations of the study*

Conclusions are limited to the effect of the factors in a learning environment designed for a single individual learning session. Participants, since not randomly selected, may not represent the whole population, suggesting that the study should be repeated to check external validity. The findings of the study are limited to one aspect of

sustainability, namely knowledge. The study did not evaluate affective or behavioural aspects.

#### *Suggestions for Future work*

The medium effect size and positive feedback from learners suggests that Systems Thinking may be useful if presented differently. It may have a stronger effect on understanding and development of transferrable skills if taught in an interactive classroom or group situation and not limited to a single session. Care should be taken not to cause excessive cognitive load on learners.

Furthermore, effect sizes in educational research are often categorised as small by Cohen's standards [66]. This is because there are typically many other important factors affecting results, typically prior education and skills such as numeracy, literacy, science and so on. Interpreting effect sizes in educational interventions is a complex matter, and evolving [*ibid*]. Where effect sizes are modest, the sample size must be increased in order to increase the power of the study. Increasing statistical power decreases the the probability of a Type II error, in which the researcher wrongly concludes that there is no effect when one actually exists [60].

Since Systems Thinking and simulation were the factors under investigation, and sustainability understanding was the dependent variable, an improvement to a future study design would be to exclude people already knowledgeable in those areas.

The methodology (factorial study design coupled with Factorial ANOVA significance testing) could be usefully employed for further studies to investigate various styles of learning intervention, such as multiple learning sessions, role play and group model building, or it could be used to evaluate existing virtual worlds, games and simulators. The effect of other factors (independent variables) could be evaluated using a similar factorial design.

## Funding Information

This research was undertaken for the PhD studies of the corresponding author at the National University of Ireland Galway (NUIG) and was supported by funding from ResponSEable (EU Horizon 2020 project number 652643), Ireland's Higher Education Authority (through the IT Investment Fund and ComputerDISC, and the Covid-19 Costed Extension), and the NUIG PhD Write-Up Bursary.

## Author Contributions

Conceptualization, C.G.; Methodology, C.G.; Software, C.G.; Validation, C.G.; Formal Analysis, C.G.; Investigation, C.G.; Resources, J.D. and O.M.; Data Curation, C.G.; Writing – Original Draft Preparation, C.G.; Writing – Review & Editing, O.M. and J.D.; Visualization, C.G.; Supervision, J.D.; Project Administration, J.D.; Funding Acquisition, J.D. and O.M.. All authors have read and agreed to the published version of the manuscript.

## Conflict of Interest

The authors declare no conflict of interest.

## Research Ethics Statements

All subjects were adults aged over 18 and gave their informed consent for inclusion before they participated in the study. For this non-interventional study, all participants were fully informed that confidentiality was assured, that data would be anonymised before publishing, why the research was being conducted, how their data would be used, any risks associated, and their right to withdraw.

## Data Availability Statement:

Data available in a publicly accessible repository. The anonymised data presented in this study, together with the R Scripts necessary to reproduce all results, are openly available in Zenodo.

DOI: <https://doi.org/10.5281/zenodo.5569508>

URL: <https://zenodo.org/record/5569508>.

## References

1. Purvis, B.; Mao, Y.; Robinson, D. Three pillars of sustainability: in search of conceptual origins. *Sustainability Sci.* **2018**, *14*, 681-695.
2. World Commission on Environment and Development. *Our common future*; Oxford University Press: 1987.
3. Rieckmann, M.; Mindt, L.; Gardiner, S. *Education for Sustainable Development Goals. Learning Objectives*; UNESCO: 2017.
4. Hardin, R.; Bhargava, A.; Bothner, C.; Browne, K.; Kusano, S.; Golrokhan, A.; Wright, M.; Zeng, P.Z.; Agrawal, A. Towards a revolution in sustainability education: Vision, architecture, and assessment in a case-based approach. *World Dev. Perspect.* **2016**, *1*, 58-63, doi:10.1016/j.wdp.2016.05.006.
5. O'Flaherty, J.; Liddy, M. The impact of development education and education for sustainable development interventions: a synthesis of the research. *Environ. Educ. Res.* **2018**, *24*, 1031-1049, doi:10.1080/13504622.2017.1392484.
6. Hofman-Bergholm, M. Could Education for Sustainable Development Benefit from a Systems Thinking Approach? *Systems* **2018**, *6*, doi:10.3390/systems6040043.
7. Hofman-Bergholm, M. Changes in Thoughts and Actions as Requirements for a Sustainable Future: A Review of Recent Research on the Finnish Educational System and Sustainable Development. *J. Teach. Educ. Sustain.* **2018**, *20*, 19-30, doi:10.2478/jtes-2018-0012.
8. Sterman, J. *Business dynamics : systems thinking and modeling for a complex world*; Boston : Irwin/McGraw-Hill: Boston, 2000.
9. Tran, L.U.; Payne, D.L.; Whitley, L. Research on learning and teaching ocean and aquatic sciences. *NMEA Special Report* **2010**, *3*, 22-26.
10. Brennan, C.; Ashley, M.; Molloy, O. A System Dynamics Approach to Increasing Ocean Literacy. *Front. Mar. Sci.* **2019**, *6*, doi:10.3389/fmars.2019.00360.
11. Frisk, E.; Larson, K.L. Educating for sustainability: Competencies & practices for transformative action. *J. Sustain. Educ.* **2011**, *2*, 1-20.
12. Wiek, A.; Withycombe, L.; Redman, C.L. Key competencies in sustainability: a reference framework for academic program development. *Sust. Sci.* **2011**, *6*, 203-218.
13. Soderquist, C.; Overakker, S. Education for sustainable development: a systems thinking approach. *Global Environ. Res.* **2010**, *14*, 193-202.
14. Cavana, R.Y.; Forgie, V.E. Overview and Insights from 'Systems Education for a Sustainable Planet'. *Systems* **2018**, *6*, 5, doi:10.3390/systems6010005.
15. Meadows, D.H.; Meadows, D.L.; Randers, J.; Behrens, W.W. *The Limits to Growth: a Report for the Club of Rome's Project on the Predicament of Mankind*; Universe Books: New York, 1972.
16. Ford, A. *Modeling the environment*, 2nd ed. ed.; Washington, DC : Island Press: Washington, DC, 2010.
17. Bossel, H. *System Zoo 2 Simulation Models: Climate, Ecosystems, Resources*; BoD—Books on Demand: 2007; Volume 2.
18. Stave, K. Participatory system dynamics modeling for sustainable environmental management: Observations from four cases. *Sustainability* **2010**, *2*, 2762-2784.
19. Martins, J.H.; Camanho, A.S.; Oliveira, M.M.; Gaspar, M.B. A system dynamics model to support the management of artisanal dredge fisheries in the south coast of Portugal. *Int. Trans. Oper. Res.* **2015**, *22*, 611-634, doi:10.1111/itor.12090.
20. Dudley, R.G. A basis for understanding fishery management dynamics. *Syst. Dynam. Rev.* **2008**, *24*, 1-29, doi:10.1002/sdr.392.
21. Sterman, J.; Fiddaman, T.; Franck, T.; Jones, A.; McCauley, S.; Rice, P.; Sawin, E.; Siegel, L. Climate interactive: the C - ROADS climate policy model. *Syst. Dynam. Rev.* **2012**, *28*, 295-305, doi:10.1002/sdr.1474.

22. Saeed, K. *Towards sustainable development: Essays on system analysis of national policy*; Routledge: 2019.
23. Moxnes, E. Misperceptions of basic dynamics: the case of renewable resource management. *Syst. Dynam. Rev.* **2004**, *20*, 139-162, doi:10.1002/sdr.289.
24. Fisher, D.M. “Everybody thinking differently”: K-12 is a leverage point. *Syst. Dynam. Rev.* **2011**, *27*, 394-411, doi:10.1002/sdr.473.
25. Sterman, J. Interactive web - based simulations for strategy and sustainability: The MIT Sloan LearningEdge management flight simulators, Part I. *Syst. Dynam. Rev.* **2014**, *30*, 89-121, doi:10.1002/sdr.1513.
26. Deegan, M.; Stave, K.; MacDonald, R.; Andersen, D.; Ku, M.; Rich, E. Simulation-Based Learning Environments to Teach Complexity: The Missing Link in Teaching Sustainable Public Management. *Systems* **2014**, *2*, 217.
27. Pallant, A.; Lee, H.-S. Teaching Sustainability Through System Dynamics: Exploring Stocks and Flows Embedded in Dynamic Computer Models of an Agricultural Land Management System. *J. Geosci. Educ.* **2017**, *65*, 146-157, doi:10.5408/16-169.1.
28. Kopainsky, B.; Sawicka, A. Simulator-supported descriptions of complex dynamic problems: experimental results on task performance and system understanding. *Syst. Dynam. Rev.* **2011**, *27*, 142-172, doi:10.1002/sdr.445.
29. Anderson, P.H.; Lawton, L. Simulation exercises and problem based learning: Is there a fit? In Proceedings of the Developments in Business Simulation and Experiential Learning: Proceedings of the Annual ABSEL conference, 2004.
30. Stave, K.; Hopper, M. What constitutes systems thinking? A proposed taxonomy. In Proceedings of the 25th International Conference of the System Dynamics Society, 2007.
31. Richmond, B. Systems thinking: Critical thinking skills for the 1990s and beyond. *Syst. Dynam. Rev.* **1993**, *9*, 113-133, doi:doi:10.1002/sdr.4260090203.
32. Forrester, J.W. System dynamics - A personal view of the first fifty years. *Syst. Dynam. Rev.* **2007**, *23*, 345-358, doi:10.1002/sdr.382.
33. Senge, P.M. *The Fifth Discipline: The Art and Practice of the Learning Organization*, Reissue. ed.; UK Century: London, 1993.
34. Richmond, B. The thinking in systems thinking: eight critical skills. *Tracing Connections: Voices of Systems Thinkers* **2010**, 3-21.
35. Kim, D.H. *Systems archetypes I: diagnosing systemic issues and designing high-leverage interventions*; Pegasus Communications: 1993.
36. Kim, D.H. *Systems archetypes II: using systems archetypes to take effective action*; Pegasus Communications: 1994; Volume 2.
37. Kim, D.H. *Systems archetypes III: Understanding patterns of behavior and delay*; Pegasus Communications: 2000.
38. Landriscina, F. Simulation-Based Learning: Open questions and guidelines for the instructionally effective use of simulation. *Journal per la formazione in rete* **2013**, *13*, 68-76.
39. Yasarcan, H. Improving understanding, learning, and performances of novices in dynamic managerial simulation games: A Gradual - Increase - In - Complexity Approach. *Complexity* **2010**, *15*, 31-42.
40. Smetana, L.K.; Bell, R.L. Computer Simulations to Support Science Instruction and Learning: A critical review of the literature. *Int. J. Sci. Educ.* **2012**, *34*, 1337-1370, doi:10.1080/09500693.2011.605182.
41. Cannon-Bowers, J.A.; Bowers, C.A. Synthetic learning environments. In *Handbook of Research on Educational Communications and Technology*, Third Edition ed.; Spector, J.M., Ed.; Taylor & Francis Group: 2008.
42. Ghaffarzadegan, N.; Lyneis, J.; Richardson, G.P. How small system dynamics models can help the public policy process. *Syst. Dynam. Rev.* **2011**, *27*, 22-44, doi:10.1002/sdr.442.

43. Reimann, P.; Thompson, K. Ecosystem modeling for environmental education: from stocks and flows to behavior and interactions. In *Model-Based Approaches to Learning*; Brill Sense: 2009; pp. 111-148.
44. Gobert, J.D.; Buckley, B.C. Introduction to model-based teaching and learning in science education. *Int. J. Sci. Educ.* **2000**, *22*, 891-894.
45. Brennan, C.; Molloy, O. A system dynamics approach to sustainability education. *Syst. Res. Behav. Sci.* **2020**, *37*, 875-879, doi:<https://doi.org/10.1002/sres.2755>.
46. Cohen, L. *Research methods in education*, Eighth edition. ed.; London, New York : Routledge: 2018.
47. Breierova, L. Generic structures: Overshoot and collapse. *Cambridge, Massachusetts, USA: Massachusetts Institute of Technology, [online]*, <https://ocw.mit.edu/courses/sloan-school-of-management/15-988-system-dynamics-self-study-fall-1998-spring-1999/readings/generic3.pdf> 1997.
48. Harris, M. *Lament for an ocean: the collapse of the Atlantic cod fishery*; McClelland & Stewart: 2013.
49. Kim, D.H. *Introduction to systems thinking*; Pegasus Communications Waltham, MA: 1999; Volume 16.
50. Arnold, R.D.; Wade, J.P. A Complete Set of Systems Thinking Skills. *INSIGHT* **2017**, *20*, 9-17.
51. Meadows, D.H. *Thinking in Systems: A Primer*; Chelsea Green Publishing: 2008.
52. Daly, H.E. Toward some operational principles of sustainable development. *Ecol. Econ.* **1990**, *2*, 1-6, doi:[https://doi.org/10.1016/0921-8009\(90\)90010-R](https://doi.org/10.1016/0921-8009(90)90010-R).
53. Banister, D.; Holden, E.; Langhelle, O.; Linnerud, K.; Meadowcroft, J.; Gilpin, G. *What Next for Sustainable Development? Our Common Future at Thirty*; Edward Elgar: 2019; p. 295.
54. Meadows, D.H.; Meadows, D.L.; Randers, J. *Beyond the limits: global collapse or a sustainable future*; Earthscan Publications Ltd.: 1992.
55. United Nations Educational; Scientific and Cultural Organization. The education for sustainable development sourcebook. **2012**.
56. Bhattacherjee, A. Social Science Research: Principles, Methods, and Practices. **2012**.
57. Johnson, R.L. *Assessing Performance Designing, Scoring, and Validating Performance Tasks*; New York : Guilford Publications: New York, 2008.
58. Sauder, D.C.; DeMars, C.E. An Updated Recommendation for Multiple Comparisons. *Advances in Methods and Practices in Psychological Science* **2019**, *2*, 26-44, doi:[10.1177/2515245918808784](https://doi.org/10.1177/2515245918808784).
59. Navarro, D. *Learning statistics with R: A tutorial for psychology students and other beginners: Version 0.5*; University of Adelaide Adelaide, Australia: 2013.
60. Sullivan, G.M.; Feinn, R. Using Effect Size-or Why the P Value Is Not Enough. *J. Grad. Med. Educ.* **2012**, *4*, 279-282, doi:[10.4300/JGME-D-12-00156.1](https://doi.org/10.4300/JGME-D-12-00156.1).
61. Cohen, J. *Statistical power analysis for the behavioral sciences*, 2nd ed. ed.; Hillsdale, N.J.: 1988.
62. Cumming, G.; Calin-Jageman, R. *Introduction to the new statistics: Estimation, open science, and beyond*; Routledge: 2016.
63. Pourhoseingholi, M.A.; Baghestani, A.; Vahedi, M. How to control confounding effects by statistical analysis. *Gastroenterology and hepatology from bed to bench* **2012**, *5*, 79-83.
64. Sweller, J. Cognitive Load During Problem Solving: Effects on Learning. *Cognit. Sci.* **1988**, *12*, 257-285, doi:[https://doi.org/10.1207/s15516709cog1202\\_4](https://doi.org/10.1207/s15516709cog1202_4).
65. Olaya, C. Cows, agency, and the significance of operational thinking. *Syst. Dynam. Rev.* **2015**, *31*, 183-219, doi:[10.1002/sdr.1547](https://doi.org/10.1002/sdr.1547).
66. Kraft, M.A. Interpreting effect sizes of education interventions. *Educational Researcher* **2020**, *49*, 241-253.