Productive Failure and Learning about Epidemics

and Complex Systems in Medical Education

Raeda Zahran Al Hinai

A thesis submitted to fulfil requirements for the degree of

DOCTOR OF PHILOSOPHY

Faculty of Arts and Social Sciences

School of Education and Social Work

the University of Sydney

September 2025

Statement of Originality

This is to certify that to the best of my knowledge; the content of this thesis is my own work. This thesis has not been submitted for any other degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis as well as all sources have been acknowledged.

Name: Raeda Zahran Al Hinai

Abstract

Over the past decades, diseases such as acquired immunodeficiency syndrome (AIDS), severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), and COVID-19 have profoundly impacted the social and economic well-being of societies. Simultaneously, the study of epidemiology has been on the rise, driven by the increased availability of data on mass outbreaks through the Internet and social media networks. These platforms have enabled researchers to monitor and analyse outbreak patterns with remarkable accuracy and efficiency, leading to significant advances in epidemiological research. In modern epidemiology education, however, educators still commonly apply what are arguably outdated teacher-centred pedagogies that focus on measuring students' capacity to memorize rather than directly assessing their explanatory grasp of the material (i.e., explanatory knowledge).

Moreover, many epidemiology concepts are operationalised as "complex systems," where the interaction of multiple distributed components generates emergent qualities of the larger population (e.g., feedback loops) not exhibited by the individual components. With many students left struggling to understand the connections between these emergent system behaviours and the underlying individual-level interactions, there is a need for novel student-centred approaches to teaching epidemiology. One promising such approach is Productive Failure (PF), where students are first allowed to grapple with problems before receiving instruction to clarify and consolidate their conceptual understanding. This contrasts to traditional Direct Instruction (DI), where instruction precedes practice and problem solving. Multiple studies across a range of natural and social sciences suggest that PF leads to greater learning gains than DI, including gains on complex systems concepts and improved ability to transfer knowledge to new domains. Compatible with PF as an instructional technique is the use of computer models, specifically Agent-Based Models (ABMs), to simulate aggregate

behaviour through the programming of interactions between individual agents. ABMs have been shown to help students visualize how complex systems phenomena unfold in epidemiology. To date, however, neither PF nor its potential implementation in conjunction with ABMs has been researched in epidemiology pedagogy.

In this quasi-experimental study of 35 undergraduate medical students, an experimental Productive Failure (PF) group (n=20) and a control Direct Instruction (DI) group (n=15) were presented with identical challenge problems via Agent-Based Models (ABMs) over the course of three two-hour online sessions, with related instruction presented either following or preceding the problems (PF or DI group, respectively). Pre- and post-tests were completed using the Moodle platform to assess the students' declarative and explanatory knowledge of complex systems in the context of epidemiology. An analysis of students' performance on the challenge problems themselves was also conducted to determine how the PF vs. DI condition impacted students' ability to transfer the knowledge gained to near or far domains. Subsequent online focus group interviews were conducted to investigate students' overall learning experience.

The results revealed no statistically significant post-test differences between the PF and DI groups in declarative or explanatory knowledge of complex systems in epidemiology. However, students in the PF group demonstrated significant improvement between the pretest to post-test on near transfer, indicating that PF may be effective in enhancing the ability to apply knowledge within instructed domains. In contrast, the DI group showed no significant gains in near transfer scores. For far transfer, the PF group significantly outperformed the DI group in its ability to apply complex systems knowledge to novel, uninstructed domains. These findings suggest that while both instructional methods may foster the development of declarative and explanatory knowledge, the PF approach holds particular promise for improving students' capacity for transfer of learning to near within and far across domains.

Future research should explore these differences with larger samples and examine the application of PF and ABMs across various medical and real-world contexts.

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my supervisors, Prof. Michael Jacobson and Associate Prof. Lina Markauskaite, for their invaluable advice, continuous support, and patience during my PhD study. Despite all the challenges I struggled with at every stage of the research project, Michael taught me that nothing is impossible and that being rigorous about solving problems is the secret of being an innovative researcher. Additionally, I am deeply grateful to Dr. Gina Arena for her treasured support, which was influential in shaping my experimental methods, and for being one of the raters. I could not have completed this thesis without her expertise in the field of epidemiology and unconditional support and encouragement.

I would like to express my sincere gratitude to the Faculty of Medicine at Sultan Qaboos University, Oman, for granting approval to conduct my study with medical undergraduates. I am especially thankful to Dr. Hana Al Sumri for her invaluable support in facilitating the study's implementation with the students. I would like to express my heartfelt gratitude to Safiya Al Shidhani, an instructional developer at Sultan Qaboos University, for her invaluable support in voluntarily designing the study content materials on the Moodle platform, diligently following up with participants, and compiling the data reports. I also extend my special thanks to, Mohammed Al Hinai, a medical practitioner, for his unwavering support in helping me understand the context of epidemics and design computational models.

I would also like to acknowledge the funding support given by, firstly, the Directorate General of Scholarships at the Ministry of Higher Education, Research & Innovation, Oman; and, secondly, the University of Sydney for offering the Faculty of Art and Social Sciences Completion Scholarship. Both sources are greatly thanked for the financial support and the facilities.

My appreciation goes out to my husband, my 'second hand', who believed in my dream and kept pushing me to proceed on my journey when I was about to give up with all the obstacles surrounding me. I cannot forget the moments when I felt that it was more important for me to live my motherhood with my six children than my dreams, as they were more deserving of that since they were still very young. Without their tremendous understanding, I would not be able to see myself reaching the end.

I am very grateful to my parents and siblings for their patience and support. This is a task that I could not have accomplished without the support from my amazing family.

Finally, I would like to express my sincerest appreciation to a number of consultants who assisted me on various aspects of this project. Felipe Haro, a simulations consultant, helped me design some of the interface of the ABMs. David Muscatello, an Associate Professor in infectious diseases epidemiology at UNSW, Australia, assisted with the selection of epidemics concepts to be taught. Mikhail Prokopenko, the Director of the Centre for Complex Systems at the University of Sydney, Australia, helped ensure that each pair of ABMs used in the sessions shared crucial structural parallels. Regarding some aspects of the quantitative analysis, I received useful input from various colleagues, including Alexandra (Ali) Green, a statistical consultant at Sydney Informatics Hub at USYD. In regard to editorial assistance, I benefited from the insights of both Autumn, an academic editor and dissertation coach, and Robert Hamilton, a researcher and editor/proofreader. For the sake of thoroughness, I might also mention that I benefited at times from the use of Consensus, an AI-powered academic search engine, when seeking sources relevant to my work.

My sincerest thanks to anyone else whom I may have missed.

Abbreviations

ABM	Agent-Based Model
AC	Analogical Comparison
DES	Discrete Event Simulation
DI	Direct Instruction
PBL	Problem-Based Learning
PF	Productive Failure
RSMs	Representations and Solution Methods
SD	System Dynamics
SEIR	Susceptible, Exposed, Infectious, Recovered/Removed
SIR	Susceptible, Infectious, Recovered

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

Epidemiology: A Complex Field

The daily impact of disease on the social and economic well-being of society is undeniable. Over the past few decades, diseases such as acquired immunodeficiency syndrome (AIDS), avian flu, severe acute respiratory syndrome (SARS), Middle East respiratory syndrome (MERS), Ebola virus, Zika virus, and others have killed thousands of individuals (Reperant & Osterhaus, 2017; Wilder-Smith et al., 2017).

Epidemiology is "the study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems" (Last, 2001, p. 61). An inherently complex field, epidemiology's concepts draw upon different definitions of populations and sub-populations presented at various levels of abstraction (Getz et al., 2015; Jacobson et al., 2016). So, for example, a population of study might be defined geographically (e.g., "all residents of county X"), occupationally (e.g., "all factory workers"), or on the basis of a specific diagnosis (e.g., "all individuals who tested positive for COVID-19"), and these populations may be mutually exclusive, overlap, or stand in subset relation (e.g., "all factory workers who've tested positive for COVID-19 in county X"; Coggon et al., 2024). Much of epidemiology's complexity arises from the fact that its central ideas are based on complex systems, a term referring to any system which consists of multiple distributed components whose interaction generates complex behaviours (Mitchell, 2009) or, more specifically, the "emergence" of "novel characteristics exhibited on the level of the whole ensemble, but not by the components in isolation" (Laszlo & Krippner, 1998, p. 68). Examples of complex systems include the flocking behaviours of birds, urban traffic patterns, stock markets, social media networks, and environmental ecosystems (Crooks &

Hailegiorgis, 2013; Drożdż et al., 2021; Goldstone & Wilensky, 2008; Jacobson & Wilensky, 2006). Each bird in a flock, for example, coordinates its movements with only a fixed number of neighbouring birds, yet the flock as a whole is able to maintain a remarkable level of cohesion, even when a predator intrudes (Feder, 2007).

The populations through which diseases spread similarly function as complex systems (Naimi, 2019; Page & Zelner, 2020) and considering them in system terms fosters an understanding of the dynamic, complex behaviours of the viruses (Hirsch et al., 2023; Xia et al., 2017). Regarding the populations in view, a core activity in epidemiology is "population thinking", which involves dividing data by different population definitions and comparing them (Gange, 2008). For example, the vectors of disease spread vary depending on whether one is examining only one individual (i.e., from the point of entry into the body, then to other organ systems) or many millions of individuals (e.g., disease spread via transportation routes). Keyes and Galea (2014) acknowledge that population-level thinking "is not easy" (p. 663), such as when categorising populations in terms of different types of individuals (e.g., neighbourhood vs. city; elderly vs. young adults; cf. Gebbie et al., 2003)

To complicate matters further, being an interdisciplinary science, epidemiology presents challenges in terminology by sometimes employing terms that are also used in other fields but with slightly different meaning (James et al., 2006; Keyes & Galea, 2014). For example, the term "population" in epidemiology typically refers to a human population of individuals, whereas in statistics it may routinely refer to any group of objects under consideration, human or otherwise, real or hypothetical (Krieger, 2012).

The Challenges of Epidemiology Education

Its complexity notwithstanding, the study of epidemiology is on the rise. Internet communication's emergence in the late 20th century, and particularly the phenomenon of social media networks, has made mass outbreaks of infectious diseases more traceable and the data on them more plentiful (Chou et al., 2018; Kouzy et al., 2020; Roberts et al., 2017). This advent of accessible data has made epidemiology an essential part of modern health and public health education curricula (Goldmann et al., 2018; Keyes & Galea, 2014). Epidemiology is typically a requirement for students in the health sciences, in allied fields like public health and health policy, and in professional schools like medicine and nursing (Goldmann et al., 2018; Keyes & Galea, 2014; Kolchraiber et al., 2019).

Not only has modern epidemiology instruction become more prominent, it has also become standardised. As reported in a 2014 article by American researchers Keyes and Galea (2014) based on an analysis of 14 introductory epidemiology textbooks published between 2002 and 2011, several topics, in particular, are considered essential to epidemiology studies. First, metrics of disease occurrence are described, such as incidence rates and level of risk, as well as methods of organising data for descriptive epidemiology. Next, descriptions and comparisons of epidemiological study designs, such as cohort studies and randomised trials, are included. Lastly, the textbooks cover statistical analysis concepts such as correlation and confounding variables. Textbooks generally have additional sections on specialised topics, but all of them present these three core areas of knowledge (Gange, 2008; Goldmann et al., 2018; Keyes & Galea, 2014).

Traditional Approaches to Epidemiology Instruction

Given the increasingly prominent role epidemiology plays in the health sciences and the challenge posed to students by the inherent complexity of the material, it is imperative that the teaching methodologies used by epidemiology instructors be effective. In modern epidemiology education, educators still commonly apply teacher-centered pedagogies through lectures and the administration of tests—methods that can be understood as assessing students' capacity to memorise more so than their deeper grasp of the concepts involved (Keys & Galea, 2014). It may come as no surprise, then, that—given epidemiology's inherent complexity as described above—students often struggle to understand the connections between health phenomena that occur at different levels of abstraction (Getz et al., 2015). Because of this, many educators are keen on developing new approaches to teaching epidemiology that would enhance students' ability to grasp and apply complex epidemics knowledge.

Productive Failure (PF) Approach

One promising approach first developed in the field of mathematics education is Productive Failure (PF), proposed by Kapur (2008) as an alternative to traditional "direct instruction" (DI). With the latter, DI, an instructional intervention in which the teacher explains the concepts that are the focus of the lesson is followed by exercises meant to give the students practice with the concept(s) in which they have just been instructed. In contrast, with the PF approach, students are first given opportunities to explore a new concept through exercises or problems with which they are likely to (at least to some extent) fail to achieve the "correct" answer. Only then, after they have wrestled with the material and developed a provisional understanding of the problem(s) involved, are they provided an instructional lesson to clarify the crucial concept(s). In a 2023 talk, Kapur used vaccinations as an analogy to his PF approach, as they share the same principle of learning: exposure to failure before achieving mastery. A vaccination trains and strengthens the immune system by providing an opportunity for it to first deal with a harmless version of a pathogen prior to combatting the actual virus in the future. Similarly, PF engages learners in complex problems before they receive instruction, resulting in a more profound understanding of the material.

The PF approach also closely aligns with the concept of preparation for future learning (PFL) discussed by Schwartz et al. (2011). PFL emphasises the importance of designing instructional experiences that enable students not only to achieve immediate learning outcomes but also to be better prepared for future instruction. Like PF, PFL values the role of initial exploration or problem-solving before instruction, as it activates learners' prior knowledge and deepens their subsequent understanding. Unlike PF, however, PFL need not focus directly on learners' struggles or failures (Schwartz et al., 2011).

Kapur's core idea is that failure can lead to greater success, specifically when it is structured and supported. Kapur argues that this approach to instruction leads to stronger and more personalised comprehension of concepts, a claim which has been supported by a number of studies both in mathematics (Kapur, 2010, 2014; Kapur & Bielaczyc, 2012) and other content fields of education such as climate science (Jacobson et al., 2017, 2020), engineering (Loibl et al., 2017), computer science (Steinhorst et al., 2024), the social sciences (Nachtigall et al., 2020), and medical and health education (Palominos et al., 2022; Steenhof et al., 2019). In Steenhof et. al. (2019), for example, the researchers found that first-year pharmacy students were better prepared for future problem solving by PF than DI.

PF's advantage over DI in facilitating learning seems especially notable where instruction of complex systems is concerned. As observed by Jacobson et al. (2017, p. 4), previous research has shown that the PF approach can "lead to significant learning of complex systems concepts" (cf. Jacobson & Markauskaite, 2015; Portolese et al., 2016), whereas there were no known published studies "in which complexity ideas were successfully learned from the use of DI learning designs" (p. 5). One possible explanation for this, as noted by Kapur (2016), is that DI has been found to facilitate the learning of *declarative* knowledge, whereas PF seems to lead to better learning outcomes for *explanatory* knowledge. As Jacobson et al. (2017, p. 4) explain, declarative knowledge is "'knowing what' or factually oriented information" (Bransford et al., 1999) and explanatory information is "knowing how or why' information" (Coleman, 1998). More exactly, declarative knowledge features a store of associations, information, and facts: information that can be presented as a statement (ten Berge & van Hezewijk, 1999). The understanding of phenomena in declarative knowledge is thus much like listing out statements without necessarily understanding how the phenomena came to be. In contrast, explanatory knowledge allows one to construct a particular explanation (Coleman, 1998); that is, given a particular piece of declarative knowledge, a person with explanatory knowledge can articulate how the phenomenon came about. In regard to complex systems, it is plausible that, as Jacobson et al. (2017) imply, although the basic concepts of complex systems can be learned in a straightforward declarative manner, explanatory knowledge is required for students to fully grasp the ways that complex systems function.

Knowledge Transfer

Another challenge faced within epidemiology education stems from the field's interdisciplinary nature: Teaching epidemiology requires the delivery of concepts that cover different disciplines, entailing the study of public health, medicine, nursing, allied health professions, health promotion, and environmental health (Banack et al., 2021; James et al., 2006). It also touches upon a wide variety of basic sciences, such as mathematics and biochemistry (Berglund, 2015). Rather than facilitate students' ability to transfer knowledge within and across these disciplines, instructors sometimes compartmentalise topics without mentioning the developmental history of the topic or explaining relevant conceptual links. For example, Keyes and Galea (2014) point out that when students learn about case-control and cohort studies, they may easily view these study designs as completely separate methods. However, both approaches share key concepts such as the identification of relationships between exposures and outcomes. If students do not recognise these commonalities, they might focus on memorising the steps for each method (treating epidemiology as simply algorithmic case-recognition) rather than understanding the broader principles that unify them. In short, emphasising the similarities and shared ideas across different methods helps students see epidemiology as a conceptual science, not just a set of isolated techniques.

Two broad types of transfer may be impeded by compartmentalisation of topics in this way. Transfer of instructional content within a subject domain or to closely associated domains with which students are already familiar is known as *near within domain transfer*; in contrast, *far across domain transfer* refers to transfer of content to a new, separate domain with which the student is less familiar or in which the student has not yet received instruction (Jacobson et al., 2020). As a simple example of near within domain transfer, being taught that frequent handwashing will prevent the spread of disease may lead the student to infer, by extension, that frequent use of hand sanitiser will do the same. Far across domain transfer would occur where, for example, a student is taught that positive feedback loops (i.e., the amplification of an initial effect through a reinforcing cycle) exist in the physical spread of a disease as each new infection increases the likelihood of further transmissions, then from this instruction the student recognises that something similar occurs in social compliance, where adherence to behaviours such as mask-wearing or social distancing encourages others to follow suit, thereby reinforcing the overall level of compliance within a community. Far across domain transfer is a desired goal for instruction given that the principles of epidemiology being taught are often very general in nature and, if grasped, should allow students to understand a much wider breadth of phenomena.

Yet, far across domain transfer does not come easily: Past studies have shown difficulty in getting students to pick out abstract features when presented with unfamiliar cases. For example, a 2015 study by Australian and American researchers showed that social science students were much more likely to group examples of physical phenomena together by surface-level subject domain content (e.g., sorting financial phenomena together with other non-financial phenomena) than were physical science students, who tended to sort by the commonalities in the underlying causes (e.g., depicting positive feedback loops) (Goldwater & Gentner, 2015). Many strategies have been proposed in the past two decades to address the difficulty of engendering far across domain transfer. Notably for our purposes, some researchers have observed the effectiveness in this regard of the PF instructional approach described above (Jacobson et al., 2020; Kapur, 2008; Loewenstein et al., 1999).

An important corollary transfer challenge confronting epidemiology education is the disconnect some students feel between the theoretical, research-oriented world of epidemiology and the practical concerns of clinical medicine (James et al., 2006). The solid vocational orientation that characterises many students in the health sciences (e.g., clinical medicine and nursing) may create a feeling that research—the bread and butter of epidemiology—is somehow divorced from medical practice (James et al., 2006; Olshan et al., 2019). As Australian researchers James et al. put it, "It is not uncommon . . . for undergraduate students in the health sciences to view epidemiological concepts as part of the 'mysterious world of research,' rather than as tools that will be both relevant and accessible to them as practitioners" (p. 575).

Agent-Based Models (ABMs) Framework

To ensure that epidemiology students are able to apply learned knowledge to real-world public health situations, educators in contemporary epidemiology increasingly attempt to integrate more effective learning strategies that engage students in practical tasks to build their real-world experience. One teaching tool that has proven particularly adept at facilitating knowledge transfer of epidemiology involves visualisations of epidemiological phenomena through agent-based modeling (Stefano et al., 2010; Bertaglia et al., 2024; Blikstein & Wilensky, 2009; Chopra et al., 2022; Gupta et al., 2024; Hoertel et al., 2020; O'Sullivan et al., 2016; Panthakkalakath et al., 2023; Robertson et al., 2024; Wilensky & Resnick, 1999; Williams et al., 2023). Broadly speaking, Agent-Based Models (ABMs) are one type of computer modelling approach for dynamic phenomena involving multiple actors (Abar et al., 2017), where "[a]gents may represent individuals, households, governments, or any other entities of interest" and "may adapt their behaviour in response to their experiences, interactions with other agents, and interactions with their environment" (Tracy et al., 2018, p. 2). In a typical public health ABM, "[i]ndividual characteristics such as demographics, health behaviours, health conditions, and health service utilisation" interact with "community characteristics" (e.g., socioeconomic status), features of the agent's social network (e.g., health conditions among acquaintances), and "ongoing processes such as aging and movement through the environment". These various characteristics "at multiple levels and the often-bidirectional processes that connect them create a system from which population health emerges" (Tracy et al., p. 20).

Not only has the dynamic visualisation made possible by ABMs been shown to support students' knowledge transfer, these models are ideal for simulating complex systems, given that "[a] defining feature of agent-based modelling is that it allows the emergence of population-level phenomena that are greater than or different from what would be expected based only on the aggregate of individual behaviours" (Tracy et al., 2018, p. 2; cf. Hunter et al., 2017). However, the use of ABMs in epidemiology education so far has focused on the implementation of ABM tools without examining the effect of the teaching approach in which the ABM is presented (Getz et al., 2015). Some authors suggest that DI approaches are likely to inhibit the benefits that otherwise might accrue from using ABMs (Kolchraiber et al., 2019). This raises the possibility that utilising ABM tools within the context of a PF approach to instruction might lead to better results.

Aims of the Present Study

It appears, then, from the above brief survey, that there are good reasons to expect a PF instructional/learning approach to be beneficial in addressing key challenges currently facing epidemiology education. Specifically, the advantages PF has been found to offer over DI in facilitating the learning of complex systems explanatory knowledge (e.g., Jacobson et al., 2017) suggests it would be a good fit with epidemiological content, which often critically involves complex systems. PF also holds promise for addressing the transfer problem in epidemiology, given the approach's success is facilitating both near within and far across domain transfer in other fields (e.g., Jacobson et al., 2020). Moreover, in view of the track record of ABMs in promoting transfer of epidemiology concepts (e.g., Panthakkalakath et al., 2023), the merger of PF with ABMs as a practice medium for students may help students better recognise real-world applications of epidemiological concepts.

The primary aim of this study is thus to investigate whether the PF instructional/learning design offers advantages over a DI approach for students of epidemiology in view of the field's incorporation of complex systems ideas and the opportunities it presents for transfer of knowledge to related disciplines. The research is broadly warranted given the need for more effective teaching approaches to biomedical (including epidemiological) courses at the introductory and undergraduate level (Goldmann et al., 2018; Donnison et al., 2016).

Research Questions

With the above considerations in mind, the overarching research question that arises is whether a PF or DI instructional/learning design featuring ABM-based challenge problems is more effective in helping students to develop knowledge of complex systems concepts in epidemiology and to transfer or apply this knowledge to solve new problems. For purposes of the present study, this broad question can be broken down into several investigative research questions (RQs) that explore specific facets of the overall problem:

> **RQ1:** Does the PF condition lead to superior learning outcomes in *declarative* knowledge of epidemics, *declarative* knowledge of complex systems concepts in epidemiology, and *explanatory* knowledge of complex systems concepts in epidemiology, as compared to the DI condition?

This first RQ seeks to replicate within the field of epidemiology education one of the chief findings of previous studies regarding PF's advantages over a DI instructional/learning design in other fields such as mathematics and climate science (Jacobson et al., 2017, 2020; Kapur, 2010, 2014; Kapur & Bielaczyc, 2012). Notably, RQ1 directly addresses the complexity issue facing epidemiology learners and seeks to determine whether the complex systems component of epidemiological knowledge content—the element that poses a notable challenge to students—can be better addressed through a PF than DI approach.

> **RQ2**: Does the PF condition lead to superior learning outcomes in the ability to

transfer knowledge of complex systems in epidemiology to new content in near within domains and in far across domains, as compared to the DI condition?

The second RQ invites an investigation of knowledge transfer within the context of epidemiology instruction, specifically whether the choice of instructional/learning design affects the robustness of any such transfer. That is, will students in the PF and DI conditions be able to transfer their learned knowledge about complex systems in epidemiology when solving problems involving new content either within or outside the content domain of the instruction (i.e., epidemiology) provided in the study?

RQ3: How does the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts affect the learning process across multiple sessions in PF vs. DI conditions?

This third RQ complements RQ1 and RQ2 by questioning how, more specifically, an ABM-based learning process promotes learning over time when introduced according to PF sequencing (i.e., prior to instructional intervention) versus DI sequencing (i.e., following instructional intervention). This RQ explores the possibility that the PF approach may provide a better fit for ABM-based practice with targeted concepts (hence, lead to superior learning outcomes) than does DI.

RQ4: How do students in a PF versus DI condition experience the learning of complex systems simulated via ABMs in the context of epidemiology instruction?

The final RQ seeks an understanding of how the students themselves perceive or otherwise experience an ABM-equipped learning process in the context of PF versus DI. Whereas the first three RQs may be investigated through the quantitative analysis of data, this last RQ is qualitative in nature and will require qualitative analysis to answer.

Generally speaking, it is hypothesised that students exposed to a PF instructional design will perform better than those in a DI treatment condition in regard to most of the variables identified in the RQs. That is, a preliminary endorsement of directional hypotheses in most cases seems to be in order. Thus, in line with the findings of Jacobson and Markauskaite (2015), Jacobson et al. (2017), and Portolese et al. (2016), students in a PF experimental group are expected to demonstrate higher performance of explanatory knowledge (although not necessarily declarative knowledge) about the targeted complex systems and epidemics concepts than students in a DI comparison group (i.e., RQ1). Similarly, in line with earlier research that has offered positive findings on knowledge transfer results with PF learning designs (Jacobson et al., 2017; Kapur, 2014; Kapur & Bielaczyc, 2012), it is hypothesised that students in a PF experimental group will attain higher performance on assessments requiring near within and far across domain transfer of epidemics and complex systems knowledge (RQ2). It is also reasonable to expect, in line with Kolchraiber et al.'s (2019) suggestion, that students in a PF condition will gain more problem-solving benefits from ABMs than their counterparts in a DI condition (RQ3). Finally, hearing the perspectives of students themselves about their experience solving ABM-based models in PF and DI instructional contexts (RQ4) may bring to light insights into the effectiveness of this technique in the classroom. More specific formulations of the hypotheses associated with the four RQs motivating the present study will be presented in Chapter Three when the study's methodology is presented.

Significance of the Research

There are several key differences between this study and previous research on PF. First, prior studies of PF have focused on STEM subject areas at the pre-university level (e.g., Nachtigall et al., 2020). There have to date been no studies of PF in university-level epidemiology and few in other university-level courses in other fields (e.g., a study involving first-year pharmacy students is described in Steenhof et al., 2019). Second, this study differs from prior PF research because it considers PF's possible role in facilitating knowledge transfer, a phenomenon particularly relevant to epidemiology education (Jacobson et al., 2020; James et al., 2006). Third, this study will be the first to explore the relative efficacy of ABM-based learning (i.e., computer modelling) within PF versus DI instructional approaches in the context of epidemiology education. Although there is considerable research that examines various kinds of modelling, including ABM models, in predicting the prevalence of infectious diseases within a community and assisting decision-makers in early intervention (e.g., Badham et al., 2018; Miksch et al., 2019), there are no empirical studies in this field exploring the efficacy of ABM paired with a PF instructional/learning design. In this way, the present study expands the literature on ABM models, and together with the previous two points, expands the literature on PF learning designs. Finally, the contemporary social context in which this research was conducted differs from prior PF research. Specifically, widespread misunderstandings and even disinformation arose from the social milieu surrounding the COVID-19 pandemic (Roozenbeek et al., 2020); hence, it has perhaps never been more important to ensure that students of health and medical sciences have a firm grasp of epidemiology concepts. To the extent that this study provides evidence for a more effective approach to epidemiology pedagogy leading to better learning outcomes, it may have consequential practical utility for mitigating such adverse developments in future epidemics.

More broadly, knowledge transfer is widely recognised to be a key educational goal across academic disciplines, as it fosters learners' ability to apply knowledge to new contexts, whether within or beyond the targeted domain. Although the present study investigates knowledge transfer in the biomedical and health sciences, the implications of our findings may extend to broader educational questions about how instruction can promote transferable understanding across contexts (Bransford, Brown, & Cocking, 2000).

From a pedagogical perspective, the present study is significant for three main reasons. First, it broadens the range of fields for which PF's potential educational contribution has been explored; this is particularly important given that many science subjects—physics, chemistry, biology, etc.—like epidemiology can be analysed in terms of complex systems. Second, this research, if the findings do support viewing PF as a more beneficial instructional design than DI for the epidemiology classroom, may encourage more epidemiology and other biomedical science teachers to utilise PF. As a result, students may be better equipped to develop both declarative and explanatory knowledge of key concepts and to transfer this knowledge when solving complex problems in new settings, particularly when learning technologies like computational models and simulation are integrated. Third, this study contributes to epidemiology pedagogy by offering a set of novel ABMs, developed by the researcher using AnyLogic software (https://www.anylogic.com), for use in the epidemiology classroom.

Definition of Terms

Agent-Based Models (ABMs): A simulation model involving the portrayal of aggregate behaviour through the programming of interactions between individual agents (Hunter et al., 2017). By solely modelling the individual interactions, the aggregate phenomena of interest are measured as an emergent property.

Complex system: A system consisting of multiple interacting components or agents whose interactions often follow relatively simple rules but lead to emergent behaviours or patterns at the macro level. These systems are characterised by features such as nonlinearity, feedback loops, decentralised control, sensitivity to initial conditions, and self-organisation. Examples of complex systems include natural ecosystems, stock markets, online social networks, and climate systems (Jacobson et al., 2017, 2020)

Declarative knowledge: A store of associations, information, and facts; information that can be presented as a statement (ten Berge & van Hezewijk, 1999). Declarative knowledge is much like listing out statements without necessarily understanding how the phenomena they describe came to be.

Epidemiology: The study of the occurrence and distribution of health-related states or events in specified populations, including the study of the determinants influencing such states, and the application of this knowledge to control the health problems (Porta, 2008). Epidemiology examines the causes of the spread of disease at multiple levels, between two individuals or through the movement of many thousands in a population.

Explanatory knowledge: Information that allows one to construct a particular explanation (Coleman, 1998). Given a particular piece of declarative knowledge, a person with explanatory knowledge can articulate how the phenomenon came to be. For example, the statement, "Coastal areas experience two high tides and two low tides every lunar day, or 24 hours and 50 minutes" is an example of declarative knowledge. *Explanatory* knowledge of this same topic might account for the preceding declarative fact in terms of the rotation of tidal bulges around the earth caused by the gravitational pull of the moon on the earth's oceans, and so forth.

Knowledge transfer: At the individual level, "how knowledge acquired in one situation applies (or fails to apply) to another" (Singley & Anderson, 1989, p. 1). Transfer is generally classified into two types: **Near transfer** refers to the application of learned knowledge or skills to new tasks or problems that are similar in surface features to the original context (e.g., transferring one's understanding of predator-prey dynamics in one simulation to another simulation of a similar ecological model). In contrast, **far transfer** involves applying that knowledge to novel or unfamiliar contexts, often across domains (e.g., applying systems thinking developed in a biology course to one's understanding of economic or social networks) (Barnett & Ceci, 2002).

Productive Failure (PF): An approach to teaching wherein students are confronted with problems whose solution involves using the target concept, are allowed to reach an impasse, and are later given the solution. The motivation for PF is the idea that "there is efficacy in *delaying instructional structure* in order for learners to generate conceptions, representations, and understandings, even though such understandings may not be initially correct" (Kapur & Bielaczyc, 2012, p. 48). This content encounter gives rise to "desirable difficulties and productive learner activity in solving problems" (Kapur & Bielaczyc).

Outline of the Dissertation

This introductory chapter has provided background information on epidemiology and the challenges in epidemiology education, along with the research problem and questions, as well as the general aim and potential significance of the study.

In Chapter 2, epidemiological concepts and the challenges inherent in teaching this interdisciplinary field are discussed, including the significance of integrating complex systems thinking into pedagogy. Traditional and innovative instructional approaches, such as Productive Failure (PF) and Analogical Comparison (AC), are reviewed, along with how these approaches could enhance conceptual understanding and knowledge transfer. In addition, the use of computational tools such as Agent-Based Models (ABMs) to facilitate active learning and model complex phenomena is discussed.

In Chapter 3, the methodology for the present study is presented. In brief, the design of the study includes two treatment groups: (1) the PF experimental group, exposed to a PF instructional/learning approach in which students solved ABM-based tasks before receiving teacher instruction, and (2) the DI control group, exposed to a DI approach in which students received teacher-led instruction before solving the ABM-based tasks. Both groups received the same ABMs, challenge problems, and teaching content. The chapter also describes the research design, data collection methods (including pretest and post-test, focus group interviews, and self-reports), and the procedures for data analysis.

In Chapter 4, the results of the study are presented, including both quantitative and qualitative findings. The analysis considers whether differences arose between the PF and DI groups in terms of declarative and explanatory knowledge as well as near within domain and far across domain transfer capabilities. The chapter further presents the findings regarding the learning process itself in the PF versus DI condition and includes insights drawn from students' own perceptions of the instructional approaches taken, based on self-reports and focus group interviews.

Chapter 5 explores the significance of the findings for an understanding of PF's contribution to epidemiology education in the form of fostering explanatory as well as declarative knowledge and far across domain transfer as well as near within domain transfer of knowledge. The role ABMs may play in this pedagogical process is also discussed, as is how the results of the study align or diverge from the existing literature and theoretical frameworks.

In the concluding chapter, Chapter 6, a concise overview of the principal findings is provided, reiterating how they address the research questions. Other implications of the findings, particularly in the context of medical education, are considered, and the limitations of the study and directions for future research are summarised.

Chapter Two: Literature Review

As noted in the preceding chapter, many epidemiological concepts are difficult for students to learn in part because attaining satisfactory explanatory knowledge of these concepts is not possible without also grasping associated complex systems concepts. The present chapter begins with a summary of the challenges presented to learners by the content of epidemiology before considering the literature on the complex systems which, in part, make much of that content so challenging. Then, consideration is given to developments in computer modelling, particularly Agent-Based Models, which have proven to be an effective means of fostering student understanding of complex systems. Finally, the Productive Failure instructional design highlighted in Chapter One is considered in some detail, as it is a promising instructional approach in which Agent-Based Models may be embedded. The ways in which Productive Failure differs from two similar (and potentially complementary) approaches, Problem-Based Learning and Analogical Comparison, will also be addressed.

Teaching Epidemiology: Challenges

The Challenge of Multidisciplinary Integration

Epidemiology instruction requires the teacher to cover concepts that span various related disciplines, including public health, medicine, nursing, allied health professions, health promotion, and environmental health (James et al., 2006). It also touches upon a wide variety of basic sciences, such as mathematics and biochemistry (Berglund, 2015). As noted in Chapter One, this multidisciplinary aspect of epidemiology presents educators with unique challenges such as the need to avoid compartmentalisation of concepts (as when case-control and cohort studies are presented as distinct cases, thereby obscuring their crucial similarities).

Epidemiology also often involves analysing phenomena across different levels of detail or “resolution”, At a “high resolution”, individual-level factors and behaviours are considered, focusing on specific actions or decisions taken by people. At a “low resolution”, group-level or population-wide phenomena are examined, where trends and aggregate effects become visible (Diez Roux, 2002). A good example of this interaction of phenomena across different resolutions is the "positive feedback loop", which refers to a situation where the outcome of an event reinforces or amplifies the original process, creating a self-perpetuating cycle. In public health, for instance, an individual's decision can contribute to a chain reaction that intensifies the overall trend, much like the cumulative effect of a snowball rolling downhill (Abdel-Sater, 2011; Wanelik et al., 2023). Thus, an individual's choice not to wear a mask may facilitate the spread of pathogens, enabling a sick person to infect others more easily. This creates a positive feedback loop (Wanelik et al.), where increasing infections lead to more opportunities for transmission, amplifying the infection rate. Here, the 'high-resolution' behaviour of an individual (not wearing a mask) directly impacts the 'low-resolution' group-level outcome: an escalating infection rate driven by the feedback loop.

Demonstrating such interconnections between individual actions and broader population-level dynamics is a primary focus of epidemiology instruction. That is, in teaching epidemiology, the goal will be to show how phenomena at different levels of resolution impact one another and to maximise understanding of the general principles that emerge. For this to occur, students must be able to recognise what information in specific cases is truly relevant to the concept being taught and what is not. In the preceding example, the relevant generalisation to be drawn from the occurrence of a feedback loop is that past/present actions may increase (even exponentially) the possibility of future events, not merely that wearing protective equipment and getting tested for infection are important measures for preventing the spread of disease.

In the literature, observations of the latter sort that are only marginally relevant at most have been called "surface-level features", whereas the features of a phenomenon central to the concept being taught are labelled "structural" (Schwartz et al., 2011). The goal of teaching is to make structural features of phenomena salient to students while leaving it up to surface-level features to provide primarily real-world context. With this approach, students will both gain declarative knowledge (i.e., "knowing what"; Jacobson et al., 2017, p. 4)—including knowledge of the surface-level detail that provides real-world context (e.g., wearing personal protective equipment helps prevent transmission of the disease to other individuals)—but also further gain explanatory knowledge (i.e., "knowing how or why"; Jacobson et al., p. 4) of the underlying reasons for the phenomena (e.g., taking measures to prevent infection from one individual to the next limits the number of infected individuals, which impedes the formation of a positive feedback loop that would otherwise escalate the infection rate).

The Challenge of Complex Systems

The above example involving feedback loops is also an example of the challenges posed by complex systems (of which feedback loops is one common feature; Ormand, 2010) in epidemiology education. As introduced in Chapter One, a complex system is one with emergent properties—i.e., organised patterns in the whole arising from the behaviours of many individual components—that cannot be derived in a linear sense from the interaction of these parts. Complex systems may feature feedback loops, self-organisation, and other sometimes unpredictable characteristics that develop over time without following a preset path (Holland, 1992; Mitchell, 2009). Small changes in such systems can cause disproportionate effects, with the whole being greater than the sum of its parts (Anderson, 1999).

Complex systems are a pivotal concept in many fields (Albert & Barabási, 2002). In biology, for example, the human body is a complex system made up of various subsystems (e.g., the circulatory system, the nervous system, and the immune system). Homeostasis, which pertains to functions such as the regulation of temperature and the concentration of various ions in the blood, is one emergent property that arises from the interactions between these subsystems (Murray, 2007). In ecology, ecosystems comprising multitudinous interdependent species and abiotic factors (e.g., weather, soil, water) give rise to predator-prey interactions, symbiotic relationships, and other emergent properties of the system (Levin, 1998). In economics, financial markets are complex systems with characteristics such as stock prices that do not follow a linear progression but are influenced by a variety of factors, including investor behaviour, public policy, social trends, and even psychological factors like fear or optimism that affect market and investor sentiment (Soros, 2008). Many social systems also qualify as complex systems, where ideas or behaviours can spread in ways resembling contagions (Saeedian et al., 2017). Indeed, many epidemic models illustrate that developments within social networks can influence how a disease spreads on a global scale (Germann et al., 2006).

The study of complex systems plays an increasingly significant and transformative role in many fields of 21st century natural and social science (Crooks & Hailegiorgis, 2013; Goldstone & Wilensky, 2008; Jacobson & Wilensky, 2006). One reason for this is that complex systems offer "new ways of organising science according to underlying principles, not according to established disciplines such as biology, physics, chemistry, and psychology" (Goldstone & Wilensky, 2008, p. 507). From a pedagogical standpoint, Jacobson (2001) observes that complex systems concepts may be beneficial to learners because, once grasped, such concepts make it easier to construct conceptual links across otherwise seemingly unrelated scientific domains (see also Goldstone, 2006; Lemke & Sabelli, 2008). Feedback loops, for instance, are a common property of many different complex systems. For example, in ecology feedback loops help account for predator-prey dynamics, whereas in economics they are needed to accurately model supply and demand. Similarly, the complex systems concept of emergent properties applies to understanding phenomena as diverse as flocking behaviours in birds and the organisation of neurons in the brain. Seeing such parallels can facilitate transfer of insights from one domain to another and help students better understand the systemic nature of phenomena (Samon & Levy, 2017).

The growing awareness of the importance of complex systems ideas has led some to advocate that complexity should be implicitly embedded into science curricula instead of being taught as a separate subject (Jacobson et al., 2017). At the pre-university level, Next Generation Science Standards (NGSS) have asserted the importance of integrating key conceptual perspectives of complex systems into science education such that complexity concepts are considered among the cross-cutting concepts (National Research Council, 2013). However, the textbooks on which classroom teachers traditionally depend do not typically provide a deep, holistic understanding of the mechanisms of complex systems or the interactions of their elements (Koppal & Caldwell, 2004; Weiss et al., 2003). Complexity ideas are inherently challenging, and the interaction and effects of complex systems are usually invisible, dynamic, and time-delayed; nevertheless, designing effective instructional approaches has the potential to facilitate their learning process (Feltovich et al., 1993; Jacobson & Wilensky, 2006; Jacobson et al., 2017; Wilensky & Jacobson, 2014).

A promising approach to the instruction of complex systems being explored by educators at all educational levels involves creating tools that allow students to visualise and experience the main operations of these systems. One example is the BeeSign simulation software (Danish et al., 2010), which was developed to engage first- and second-grade students in learning from a complex systems perspective how honeybees collect nectar, specifically observing the different levels of the beehive system and the inter-connectedness between its elements. Using the software, students can adjust relevant variables to observe the impacts on either the honeybee hives or flowers. Research has shown that the young students' interactive engagement with the program improves their understanding of the hives' functions, micro- and macro-level behaviours, and emergent properties (Danish et al., 2011). Another simulation established by Grotzer et al. (2013), called ECOMUVE, encourages middle school students to explore a virtual ecosystem platform to understand the cause-and-effect relationships and ecosystem processes involved, resulting in an enhancement in their comprehension of this complex system. As for undergraduate students, the BioQUEST Curriculum Consortium (http://www.bioquest.org) offers a range of computational tools and supplementary materials designed to support biology instructors in teaching complex systems. Similarly, the Cell Collective platform (https://cellcollective.org) provides an online environment that enables teachers, students, and researchers in the life sciences to collaboratively simulate and analyse various biological processes. These processes include modelling cellular development and differentiation, understanding disease pathogenesis, and evaluating the effects of different treatments on diseases. Notably, this platform allows users to engage with these complex tasks without requiring specialised mathematical or programming skills (Helikar et al., 2015).

The Use of Computer Models in Epidemiology Instruction

All the above simulation techniques are precursors to or instances of a component of the instruction of complex systems that has become a 'hot' topic of research in recent years: computer modelling. The emergent complex behaviours of a group that arise from the behaviours of its individual parts are often difficult to predict a priori solely by examining the characteristics of each part (Cuevas, 2020; Miksch et al., 2019). Computer models are able to simulate the emergence of such behaviours over time, identifying root causes and enabling predictions of future developments. Indeed, such simulations may be the only source of information available when experimental findings are too costly to procure or when experimentation may encounter ethical constraints (Cuevas, 2020; Currie et al., 2020; Li, 2018). For example, some imaginable experiments with actual ecosystems in environmental science might entail damaging biodiversity or disrupting the ecological balance. In such cases, computer simulations may be a better alternative, as researchers can safely model outcomes without risking any actual harms to the environment (Williams et al., 2023).

Epidemics have the characteristics of a complex system, so their development and spread has, not surprisingly, been subjected to computer modelling, from simulations of the effects of international travel on the transmission of COVID-19 to assessments of the effectiveness of contact tracing and isolation strategies (Fang et al., 2020; Gostic et al., 2020; Muellner et al., 2018). Various methods have been developed to model epidemics, with a number of reviews of the approaches available (Bai et al, 2020; Currie et al., 2020; Narassima et al., 2020; Nicholas & Christophe, 2008; Shatnawi et al., 2013; Tako & Kotiadis, 2015). Three broad classes of computer simulation models are generally identified: System Dynamics (SD) models, Discrete Event Simulation (DES) models, and Agent-Based Models (ABM) (Currie et al., 2020; Epstein, 2009). I will consider these in turn, then summarise the advantages that the latter type (ABMs) has over the others.

System Dynamics Models

System Dynamics (SD) approaches seek to model the progression of an epidemic by describing the collective changes of specified cohorts or compartments that interact with one another (Currie et al., 2020; Dietz, 1979). For example, the seminal SIR model (Susceptible, Infectious, Recovered) categorises the population affected by an epidemic into three groups or “compartments”: those who are Susceptible to infection, those who are presently Infectious, and those who have Recovered or have been otherwise Removed from the population (e.g., by death). Using coupled differential equations, the SIR model describes the change in the number of people in these three compartments over time (Cuevas, 2020; Silva et al., 2020). The strength of SD modelling is that it allows a strategic view of the entire system (i.e., the epidemic) to be displayed through the examination of only a few variables

(Yang et al., 2011). However, it is precisely because the model recognises only a few variables that SD modelling has shortcomings in fidelity. For instance, the SIR model assumes that all individuals in a population are equally susceptible—what has been characterised by some authors as a 'perfect mixing' assumption that is, in fact, inaccurate (Mossong et al., 2008; Shatnawi et al., 2013). For example, people who travel frequently or who live in close proximity to those who do would have a higher susceptibility to infection than those who decide to stay in their homes during the observed time period.

This shortcoming can be offset to some extent by adding more compartments or variables to increase the descriptiveness and precision of SD models. For example, a compartment containing individuals who have been infected but are not yet themselves infectious to others might be added to an SIR model. Such "SEIR" (Susceptible, Exposed, Infectious, Recovered/Removed) models have been used to study epidemics in the past and were the most common SD model used to study the COVID-19 epidemic (Currie et al., 2020; Silva et al., 2020). Some research teams used SD SEIR models to study the impact of mass social isolation policies in China in response to COVID-19 (e.g., Currie et al., 2020). In one example of adding compartments, Ivorra et al. (2020) expanded the number of compartments in an SEIR model to include infectious but undetected individuals in addition to those who had been hospitalised.

In short, the simplicity of SD models is both their strength and weakness. Fewer compartments of individuals mean fewer variables to track, leading to a relatively accessible, straightforward overall picture of epidemic dynamics. However, limiting the number of variables in this way potentially underestimates the complexity of these very same dynamics. Conversely, increasing the number of variables allows for greater fidelity of the findings while at the same time potentially diminishing their accessibility.

Discrete Event Simulation Models

If a finer-grained analysis is desired, one may choose to examine pandemics at the task or event level. Here, the approach of Discrete Event Simulation (DES) models may be helpful. In DES, dynamic situations or processes are modelled as "sequences of discrete events that occur at intervals" (Akpan et al., 2024) or, put differently, as actors who queue up for services which feed into one another in a network (Currie et al., 2020). This structure immediately lends itself to models of resource availability, such as modelling hospital wait times (Vázquez-Serrano et al., 2024) or distribution of vaccines (Currie et al., 2020). In addition, DES models can simulate negative outcomes such as the spread of an infectious disease throughout a population. This might include modelling finer elements of transmission dynamics, such as including a status for individuals who are infected but not yet infectious to others. All this makes DES a powerful tool for studying epidemic phenomena. Indeed, during the COVID-19 pandemic, several researchers called for the use of DES applications to simulate how healthcare systems responded when facing varying levels of strained resources and encountering different types of public health interventions (e.g., Narassima et al., 2020).

The chief benefit of DES models is their ability to "capture a system's behaviour and interconnection effects, which result from the combinations of many random processes, coupled with the system structure" (Vázquez-Serrano et al., 2021). This focus on "processes" allows DES models to represent systems "at an operational level" and simulate changes that "occur at discrete epochs over time" (Vázquez-Serrano et al.). This approach allows researchers to study otherwise difficult-to-capture phenomena such as cascading effects on infection rates from staggered vaccine rollouts or impacts on patient outcomes from altered hospital triage protocols.

However, DES modelling comes with challenges. For one, it can be time-consuming and costly to develop a DES model, and any team developing such a model needs to possess expertise in both simulation design and knowledge of the specific domain. Moreover, the team will need access to high-quality data that adequately and accurately captures the behaviour of the system in view. Incomplete or faulty data can result in unrealistic models that are of questionable use for decision-making (Vázquez-Serrano et al., 2021).

Agent-Based Models

As noted in Chapter One, Agent-Based Models (ABMs) model interactions at the micro-level of individual agents, simulating the "simultaneous operations and interactions between autonomous agents as actors" (Akpan et al., 2024). The model then considers the emergent properties that arise at the macro-level of the larger system from these multitudinous interactions (Wilensky, 2015). Modelled phenomena consist of three components: agents (i.e., individuals or objects), an environment, and the agents' methods of interaction. Simulations depend on determination of the agents' methods of interaction, that is, the attributes influencing how they interact either with each other or with the environment (i.e., which agents interact at which times and in what way) (Wilensky & Reisman, 2006).

An ABM's modelling of each individual in the population adds a significant measure of heterogeneity to epidemic models (Currie et al., 2020; Tako & Kotiadis, 2015). Each agent is attributed a set of individual and interactive behaviours that may be chosen from a probability distribution. Some behaviours, like movement, may be explicitly time-dependent, whereas others, like disease status, may not. In the model, all individual agents are placed in an arena where they encounter one another, and time is simulated forward stepwise (Badham et al., 2018; Silva et al., 2020). The current status of each individual is thus used to determine their status in the following increment of time.

Advantages in Simulation Capabilities of ABMs Over DS and DES Models

Despite the level of detail attainable by SD and DES models, they still describe the ensemble behaviour of populations or tasks, with the number of variables typically far from the number of players acting in the population. The perfect mixing assumption of SD noted above may struggle to describe the sharp impact of, for example, a single sick person's decision to fly to a previously unexposed country to visit relatives. DES is similarly constrained: The types of services and events operate the same way for all, and the flow of actors is pre-determined via the fixed connections between events. This becomes a fidelity problem if a given individual happens to be too far away in real life to take a trip to reach a particular service. A significant degree of homogeneity is, thus, built-in to both SD and DES models—a constraint that is directly countered by the relative heterogeneity of the ABM approach (Colizza et al., 2007).

This heterogeneity of ABMs translates into a number of distinct advantages in simulation capabilities of ABMs over SD and DES models. First, regional variables can be incorporated into ABMs (Cuevas, 2020; Yang et al., 2011). While it may be difficult to figure out a way to translate choke points on highways and country roads into differential equations for an SD model, for example, narrower and wider passageways between regions can be added into an ABM by simply changing the shape of the arena. DES's fidelity problem of individuals being too far to reach a particular service could be avoided with little or no changes to an ABM since it is designed to capture the location and distance of agents relative to their environment (Chan et al., 2010; Mossong et al., 2008). Second, an ABM allows changes to individual behaviour to be made relatively simply by changing or adding the behaviour to the model of the individual and cloning that change throughout any other agents as needed (Dietz, 1979; Hunter et al., 2017; Miksch et al., 2019). This may be easier than attempting to derive a model equation for how the single behaviour would manifest itself throughout a compartment of people in an SD model or figuring out what connections to make between events in a DES model (Narassima et al., 2020). On a related note, ABMs are well-suited to simulate the sorts of unexpected behaviours the other two methods may be illsuited to capture. For example, an ABM can incorporate simulation of a situation where one infectious individual decides, after a set amount of time, to take an airplane to a region in the arena previously isolated from exposure to a pathogen.

Despite these advantages over the other modelling approaches, ABMs do face some challenges. For one, translation issues between data and code of the sort faced by SDs and DESs can still be present with ABMs. While it may be challenging to create an ensemble equation to model emergent behaviour in an SD model, for example, it may be just as difficult to create a variable in an ABM representing a behaviour such as how often a person prefers to walk instead of to drive (Badham et al., 2018). Another problem facing ABM simulations is the large number of individual agents required for some phenomena of epidemics to emerge. This may make computation times for ABM simulations onerously long, although some researchers have suggested coding practices to mitigate this (Miksch et al., 2019; Willem et al., 2015).

Learning Benefits of ABMs

Various studies have demonstrated the benefits of implementing the computational ABM approach in learning scientific concepts, as these models can promote the development of students' causal reasoning skills and their comprehension of the mechanisms underlying the phenomena (Levy and Wilensky, 2008). A sampling of these studies spans physics (Brady et al., 2015; Sengupta & Wilensky, 2009), biology (Dickes & Sengupta, 2013; Dickes et al., 2016; Kim et al., 2015; Kitano, 2002), chemistry (Holbert & Wilensky, 2014; Levy & Wilensky, 2009; Mill et al., 2016), critical engineering (Lai et al., 2017), and climate science (Jacobson et al., 2017). At the undergraduate level, Blikstein and Wilensky (2010) tasked students with using ABMs to simulate certain concepts of material science such as atomic movement, grain growth, and food chains. They found that the use of ABMs reinforced students' ability to conceptualise the unseen processes of causality in ways that went beyond mere memorisation of theoretical content (cf. Goldstone & Wilensky, 2008; Wilensky & Reisman, 2006). Hsiao et al. (2019) similarly affirmed the use of ABMs for helping students engage in mechanistic reasoning (i.e., grasping the reasons that phenomena occur the way they do) as well as problem decomposition (i.e., breaking down problems into simpler components), both being essential skills in scientific modelling (Sengupta et al., 2013). It should be noted that students themselves have reported positive impressions of the incorporation of ABMs into classroom instruction. For example, Ceberio et al.'s (2016) study of university-level physics instruction using AMBs found that students believed the ABMs allowed them to be more proactive in their learning and to more easily overcome difficulties they encountered with the material.

In regard to ABMs' potential for helping students grasp complex systems, Jacobson et al. (2006) found that incorporating ABMs into instruction about climate change and complex systems concepts, including emergent properties, dynamic equilibrium, and non-linearity, positively affected high school students' understanding of such largely invisible, dynamic, and unpredictable scientific phenomena. Silverman et al. (2021) likewise observe that ABMs have promise in the health sciences due to their ability to simulate complex systems with features such as "emergence, non-linearity, and adaptive behaviour". On a related note, previous research suggests that the cross-level representations of scientific phenomena provided by ABMs can help learners make connections between their reasoning about elements at the individual agent level with system properties at the aggregate level (Jacobson & Wilensky, 2006; Levey & Wilensky, 2008; Sengupta & Wilensky, 2009; Wilensky & Resnick, 1999). Sengupta and Wilensky (2009), for example, used the ABM approach to teach electricity concepts and reported that undergraduate students following ABM-supplemented instruction were better able to comprehend the behaviours of electric circuits and explain multi-level relationships between electric current, voltage, and resistance at the aggregate level as these arise from the interactions between individual agents.

In general, then, ABMs, with their ability to simulate how group-level phenomena may emerge from the interaction of multitudinous individual-level decisions or events, show promise as a particularly effective tool for instruction in epidemiology. Users of ABMs can visualise both levels of resolution simultaneously and draw inferences based on the interaction of the two. The principles of infection control, for example, illustrate how individual actions or decisions (e.g., the decision to not wear a mask) can affect the larger group. Conversely, group actions may affect the individual, such as when the majority of a population complies with infection testing, thereby creating peer pressure for untested individuals to likewise comply (Parker et al., 2013).

It should be noted that ABMs' ability to simulate phenomena in a software "sandbox" environment makes them a good fit for epidemiology lessons built around Productive Failure (PF) as the instructional method (see below). That is, the ABM sandbox environment allows for correct and incorrect moves alike to be made without judgment being passed by the software. This opens up the possibility of exploring complex phenomena at varying levels of detail in ways that align with the initial exploratory problem-solving phase of PF. Similarly, ABMs can 'show' learners what are often very complex interactions involving large numbers of variables and actors that may be difficult to explain in a strictly didactic manner without first being visually operationalised. PF's implement-and-learn approach to lessons pairs well with these features of ABMs.

Epidemiology Instruction, Productive Failure, and Similar Instructional Approaches

As promising as ABMs are as a vehicle for teaching complex systems concepts in epidemiology, computer simulations are more an instructional tool than a model for instructional design. That is, ABMs need to be situated within a broader teaching method or approach that will, for present purposes, align well with the demands of epidemiological content surveyed in Chapter One. As noted there, PF appears to be a strong contender in this regard. Not only has PF been shown to be an effective approach to learning complex systems concepts (Jacobson & Markauskaite, 2015; Jacobson et al., 2017; Portolese et al., 2016), it holds promise for supporting the ever elusive far across domain knowledge transfer (Jacobson et al., 2020; Kapur, 2008; Loewenstein et al., 1999; Roll et al., 2018). This final section of the chapter first briefly considers the general state of epidemiology pedagogy today and then delves more deeply into the literature surrounding PF, comparing it to its traditional competitor, Direct Instruction (DI), as well as two somewhat similar and complementary instructional design approaches, Problem-Based Learning and Analogical Comparison.

Epidemiology Instruction Today

Epidemiology education has changed dramatically over the past few decades, driven by advances in public health, improved vaccine technologies and related innovations, and changing needs on the ground. Traditionally, epidemiology pedagogy relied heavily on lecture-based formats where students were presented with theoretical concepts and statistical techniques in a manner that was relatively isolated from other fields of research. However, as the purview of epidemiology has expanded over the years, encompassing topics such as chronic disease epidemiology, environmental health, and global health challenges, epidemiology pedagogy has evolved as well (Olsen et al., 2015; Werler et al., 2019).

Today, there is a greater emphasis on integrating real-world applications into the curriculum. Epidemiology students are expected not only to grasp theoretical concepts but also to apply these concepts to practical, hands-on situations (Hossain, 2022). This shift is evident in the growing use of case-based learning, where students analyse specific public health problems, as well as in experiential learning methods, such as fieldwork and internships, where students engage with communities and health professionals (Nelson et al., 2018; Donkin et al., 2023). This evolution is also marked by a growing emphasis on interdisciplinary collaboration. Epidemiology is no longer taught in isolation but is integrated with other public health disciplines such as environmental health, social sciences, and health policy. As a result, educators are adopting more collaborative teaching methods, encouraging students to engage with professionals from different fields and bringing a broader, more holistic perspective to the study of public health issues (Earl, 2009; Gange, 2008).

Productive Failure Versus Direct Instruction

Within this developing instructional milieu, PF has attracted growing interest as a student-centred alternative to traditional instructional approaches. PF was first induced in a 2008 paper by Manu Kapur, who found that students who engaged in a period of attempted problem-solving in ill-structured tasks subsequently performed better on well-structured, modified versions of those tasks than those who solely worked on the well-structured task alone (Kapur, 2008). Kapur proposed this "productive failure" approach as an improvement over the traditional "direct instruction" (DI) approach, and in later papers Kapur discussed how to implement PF as an instructional method (see especially Kapur & Bielaczyc, 2012). There are two main phases of student engagement in the PF approach: a problem-solving or "exploration" phase in which learners "generate and explore the affordances and constraints" of representations and solution methods (RSMs) for the problem at hand, followed by an instructional or "consolidation" phase in which the instructor builds on the student's ideas while explaining the concepts relevant to the generally accepted expert problem solutions (i.e., "canonical RSMs") (Kapur, 2024). Thus, in PF, students are given the opportunity to compare and contrast their often suboptimal and even erroneous solutions in the exploration phase to the instructor information provided in the consolidation phase.

As described by Sinha and Kapur (2019), the traditional DI teaching method (with which PF is contrasted) is similarly marked by two phases. In the first phase, the instructional phase, the teacher gives instruction on the concept that is the focus of the lesson. This may involve a description of the concept in general, along with specific examples. In the next phase, the problem-solving phase, students are given practice problems to complete, the idea being that it is beneficial for them to have an immediate opportunity to apply the instructed concepts while the concepts are still held clearly in memory. Another important motivation in DI for ordering the instructional phase first comes from cognitive load theory in psychology (Sweller, 2010; Sweller et al., 1998, 2019), which holds that tasks involving more "interactive elements" (i.e., "concepts or procedures that need to be learned" and that have an "intrinsic connection . . . necessitating their simultaneous processing in working memory for comprehension"; Chen et al., 2023, p. 62) impose a higher cognitive load on the learner. This load can be made more manageable, DI proponents contend, by sequencing problem solving activities—which tend to have higher element interactivity—after instruction (Chen & Kalyuga, 2020; Kalyuga, 2015). That is, without "prior knowledge" from guided instruction, problem solvers would face a "time-consuming search through the solution space" when attempting "to engage in sensemaking via trial and error, thereby burdening the limited capacity of the working memory" (Sinha & Kapur, 2021).

Generally speaking, element interactivity, as defined by Sweller (2010), refers to the number of information elements that require simultaneous processing of multiple elements in working memory to ensure meaningful learning. When students engage with high element interactivity task—such as feedback loops or non-linear relationships in complex systems thinking—they must connect several pieces of information in working memory. While earlier views have seen this as a potential source of cognitive overload, recent research (e.g., de Jong et al., 2024) highlights that high element interactivity can enhance deeper conceptual understanding and promote transfer across domains when learning environments are well designed and scaffolded. This perspective is suited to the Productive Failure (PF) approach, as the latter allows learners to grapple with high-interactivity elements before receiving structured instruction.

Kapur challenged the DI instructional design structure, theorising that a prior period of exploration of the problem allows the learner to understand both the causes of failure as well as success, potentially leading to better outcomes during the actual problem-solving period (Kapur, 2008, 2016; Kapur & Bielaczyc, 2012). A lesson plan based on this approach would reverse the normal order of instruction and problem solving as compared to DI. The first phase of a PF lesson thus involves students attempting to solve example problems. The goal, however, is not for them to reach correct answers or make no errors as much as simply to allow them to explore the systems in question—creating their own mental model or schema of the situation (cf. Gick & Holyoak, 1983). In the second phase of a PF approach, the instruction phase, the teacher presents the solution to the problem along with the underlying theory. This leads the student to compare their own mental model with the explanation of the teacher, potentially leading to a more memorable and personal understanding of the lesson material. Indeed, as noted in Chapter One, the PF method has demonstrated considerable success in the classroom (see Sinha & Kapur, 2021, for an exhaustive review). For example, in Kapur's own studies involving hundreds of mathematics students, PF methods have led to significantly higher increases in test scores as compared to DI (Kapur, 2010, 2014; Kapur & Bielaczyc, 2012).

This does not, however, mean that DI as an approach has no utility. One proposal that carves out a legitimate space for DI as opposed to PF comes from consideration of the "element interactivity" concept in cognitive load theory discussed above (i.e., the number of concepts that must be simultaneously processed in working memory). It has been suggested (Ashman et al., 2020; Loibl et al., 2017) that the DI approach may be most helpful when the topic being learned has a particularly high degree of element interactivity. Conversely, when element interactivity is relatively low, a problem-solving-first approach such as PF has been found to be more effective (Loibl et al., 2017). Furthermore, an instruction-first approach may offer benefits when procedural rather than conceptual knowledge is the focus of a topic, according to Chen and Kalyuga (2020). Procedural knowledge often entails algorithms and methods common to tasks such as solving math equations. Conceptual knowledge is rooted in the understanding of theories, structures, and classifications, and is key in the understanding of epidemic diseases.

When it comes to epidemiology, undergraduate students with an elementary knowledge of the field (both via class instruction and knowledge disseminated in recent years in response to the COVID-19 pandemic, for example) tend to have background familiarity, albeit limited, with epidemic topics such as those presented in the current study (see Chapter Three). This familiarity with the topic material should serve to lower the element interactivity of the material. Moreover, learning epidemiology involves grasping conceptual knowledge of complex systems, as discussed earlier. Both of these factors lend support to the notion that PF may be the preferable approach to epidemiology instruction.

Although Productive Failure (PF) is often found to support deeper learning than Direct Instruction (DI)—particularly with regard to explanatory knowledge—research also indicates that high-quality instructional videos can significantly enhance learning outcomes across diverse educational settings. Well-designed videos provide structured explanations, support knowledge retention, and reduce extraneous cognitive load (Andrade et al., 2015; Brame, 2016). The standardisation and instructional clarity afforded by such materials may increase the baseline performance levels in both PF and DI conditions, especially when both groups are exposed to the same video content. As a result, it can be challenging to isolate the effects of the pedagogical sequence alone (Andrade et al., 2015; Brame, 2016).

While studies on instructional sequencing suggest that self-directed exploration followed by instruction—as in PF—can be particularly effective, DI learners may also benefit substantially from the immediate clarity offered by a strong video (Ott et al., 2024). Furthermore, Woods and his colleagues (2007; see also Ignacio, 2022) emphasise the importance of cognitive integration, the process by which learners connect new ideas to prior knowledge, particularly in complex or interdisciplinary domains. High-quality instructional content may foster this integration, thereby contributing to deeper conceptual understanding regardless of sequencing.

Productive Failure and Problem-Based Learning

One instructional approach that bears some resemblance to PF is Problem-based Learning (PBL), which is a teaching approach that uses complex, real-world problems as a vehicle for students to learn problem-solving skills. In PBL, rather than receiving direct instruction, students collaborate in small teams to collectively acquire knowledge of a problem and interactively integrate that knowledge to arrive at a solution (Duch et al., 2001). PBL and PF are similar in that both are centred on student involvement or engagement in problem-solving activities (Pascarella & Terenzini, 1991). PBL, however, occurs within the context of collaborative groups, complete with a facilitator who guides exploration and discussion of the basic elements of the real-world case study in question. Students learn the content based on group-directed exploration but must utilise self-directed reflection and reasoning to identify what is known, what needs to be known, where to access the information, and how to use this to resolve the problem (Wood, 2003).

Pourshanazari et al. (2013) found that although PBL was less effective than the traditional lecture method in the short term, this result was reversed over the long term (i.e., one year and four years after initial instruction). Similarly, another study conducted by Strobel and Barneveld (2009) found that PBL significantly promoted long-term retention of knowledge and the development of other essential skills when compared to the traditional lecture method.

While PBL and PF share the use problems or problem cases as an important component of the learning activity, there are key differences. As noted earlier, the problem solving in PF activities occurs in an exploration phase, which is then followed by a consolidation phase in which the instructor shows how the students' ideas relate to a canonical, expert solution. This instructor-directed approach is different from typical PBL collaborative and self-directed activities, in which, as Wood (2003) explains, "students use 'triggers' from the problem case or scenario to define their own learning objectives." They then carry out "independent, self-directed study before returning to the group to discuss and refine their acquired knowledge. Thus, PBL is not about problem solving per se, but rather it uses appropriate problems to increase knowledge and understanding."

Broadly speaking, there has been far greater use of the PBL instructional/learning design in the medical and health related sciences than there has been of PF (see Steenhof et al., 2019, for an example of the latter). Wood (2003) describes PBL as "common" in medical schools; indeed, the approach was pioneered in such a school in the 1960s (Barrows, 1996). Despite the differences between the approaches, at least one attempt has been made to blend the use of PBL and PF: Portolese (2021), in her thesis, integrated elements of a PF learning design into the closing phase of a PBL approach in order to examine undergraduate medical students' ability to achieve deeper learning and near within domain transfer of knowledge. In this study, all students carried out PBL-style self-directed activities; however, in the final phase of the study, an experimental group received a PF-style intervention in which tutors offered direct feedback on the rightness or wrongness of the students' work (mirroring the consolidation phase of PF). Portolese found that the experimental group, which received this PF-style intervention, overall performed better than the control group, which received only a typical "'hands-off" approach to a tutor's role in the PBL closing" (p. 100). That is, the experimental group exhibited significantly greater explanatory understanding of the concept involved as well as ability to transfer these concepts. Given that this result amounts to an enhancement of the effectiveness of PBL when supplemented with PF, Portolese's study once again confirms the utility of PF for promoting explanatory knowledge and the ability to transfer the knowledge gained.

Although Kapur (2008; Kapur & Bielaczyc, 2012) initially highlighted Productive Failure (PF) as a more effective alternative to the traditional Problem-Based Learning (PBL) approach—arguing that PF engages students in experiencing failure while exploring and generating solutions before receiving instruction—his later work offered a more nuanced perspective. Specifically, Kapur (2014) introduced the concept of *productive success*, suggesting that students can solve problems correctly from the outset when they possess sufficient prior knowledge and when the learning experience is well-scaffolded. In other words, in well-designed PBL settings, initial success—not just failure—can be productive if it fosters deep conceptual understanding and knowledge transfer. This framing contrasts with the more common narrative around PF, which emphasises the benefits of initial struggle prior to structured instruction. Kapur (2011) further argued that the effectiveness of either initial success or failure depends largely on the nature of the task, the learner's preparedness, and the instructional sequence. Clarifying this distinction situates the present study within a more refined understanding of the instructional potential of PBL, highlighting that both success and failure can lead to productive learning when appropriately designed.

Productive Failure and Analogical Comparison

Analogical Comparison (AC) is one other teaching strategy that resembles PF and that may hold promise for epidemiology instruction. First described by Gentner and colleagues in 2003, AC proponents propose that deeper "structural" features of phenomena (in contrast to the less relevant "surface" features) can be understood more readily through comparisons of simultaneously presented examples, even if learners are not familiar with the content in the subject domains of either example. This contrasts with simple analogical teaching, which presumes students' familiarity with the targeted concept in one domain before presenting it in another (Gentner et al., 2003). In a series of experiments, Gentner and colleagues demonstrated that the task of listing similarities and differences between two cases more effectively supported knowledge transfer than feature extraction of the same cases presented sequentially.

This effect has been demonstrated in several studies over the last two decades. For example, experiments conducted by Kurtz and Loewenstein (2007) found that sequential solving of variations of Duncker's Radiation problem—a trade-off problem involving the treatment of cancer—resulted in poorer outcomes than being given two variations of the problem at once and being told to construct a similar solution for both. The simultaneous comparisons used in AC have been argued to be more effective than sequential presentation partly because the novelty of both situations in the former case more readily promotes abstract thinking, whereas the latter may cause learners to become fixated on surface-level details if familiarity is established with just one example beforehand (Gentner et al., 2003; Kurtz & Loewenstein, 2007).

The fact that research supports the effectiveness of AC and PF may at first glance suggest that learning best occurs in a seemingly unguided way, where problems and examples are simply handed to learners without any further context given. AC does not require that domain-specific content be provided for either problem being compared, only that structurally similar examples are compared simultaneously (Gentner et al., 2003). Likewise, PF involves the exploration of unfamiliar contexts and even requires failures of understanding as part of the learning process. However, there are reasons to expect that unsupervised learning in its pure form is not particularly effective. As noted by Sinha and Kapur (2019), PF is expected to fail when no feedback is given on the performance of the students during the instruction phase, as students may then have little idea of which factors were relevant to their demise and which were not (Sinha & Kapur, 2019). Indeed, some authors have brought up the possibility of inefficiency in learning using the PF system if students are unable to recognise their failures during the exploration phase (Sinha & Kapur, 2021b; Kirschner et al., 2006). Kapur recognises this potential pitfall and calls it "unproductive failure"—when the student gains no intuition of the relevant factors in the problem during exploration (Kapur, 2016). In fact, a review by Kirschner and colleagues showed that minimally guided learning is significantly less effective than even the DI teaching method (Kirschner et al., 2006). PF properly executed, however, is not an example of such "unguided learning": The instruction phase of PF is included after the exploration phase precisely so that the factors which lead to success and those which lead to failure—as well as which phenomena in the problem were relevant to the goals of the lesson—may be pointed out.

Follow-up instruction has been shown to enhance the learning effect of AC as well: A study by Stanford researchers Schwartz and colleagues found that a period of instruction after students did an AC activity improved learning outcomes (Schwartz et al., 2011). Along the same lines, a meta-analysis by Alfieri et al. (2013) concluded that knowledge transfer was facilitated when an analogical case comparison was followed by explicit presentation of the principle or principles (i.e., structural features) shared by the cases. When carried out in the context of a PF approach, this explicit presentation would normally involve a discussion of the cases as part of the instruction featured in the consolidation phase (cf. also Cao, 2020).

At least one study has emerged exploring the use of the three concepts PF, AC, and ABMs (discussed earlier) to teach topics involving complex systems. Jacobsen et al. (2020) examined whether requiring students to work with two ABM simulations in different domains of knowledge (climate science and another non-climate-science domain) would yield better learning outcomes than having them work with two ABMs from the same domain (two from climate science). The lesson was structured in a PF format: The two-model group was exposed to the PF-AC method of learning, while the one-model group solely used PF. Over four days, both groups did activities and were later quizzed on long-form tests requiring them to explain concepts from complex systems science, including feedback, chaos, and tipping points. They were also asked to solve a problem in an unrelated subject domain that would require the use of complex systems concepts, a test of far across transfer ability. The results showed that there was no significant difference between the one-model and two-model groups in their ability to explain complex systems concepts abstractly; however, there was a significant improvement in the two-model group in the ability to solve the far across transfer question (Jacobson et al., 2020). Although only a single study, Jacobson et al. (2020) suggests that a PF approach that makes use of AC and incorporates ABMs into the problem-solving elements of the design can aid in the learning of complex systems concepts and boost the level of comprehension required for far across transfer of knowledge.

The potential of merging ABMs and AC into an overall PF instructional design is bolstered as well by Goldstone and Wilensky's (2008) argument for the crucial role in facilitating knowledge transfer played by what they term "grounded generalizations", or idealised "situation construals that are concrete insofar as they are perceptually, temporally, and spatially grounded" (p. 466). In the case of complex systems, such grounded generalisations are, in effect, the insight a learner has into the deeper, structural features of the system. Importantly, Goldstone and Wilensky argue that simulation modelling approaches such as ABMs offer a particularly rich conceptual/structural environment in which such grounded generalisation can be developed. That is, working with ABMs requires learners to make interpretive choices that foster their ability to perceive similarities at the formal, structural level between different systems in different domains. This, in turn, directly facilitates transfer "not by applying a rule from one domain to a new domain but rather by allowing two scenarios to be seen as embodying the same principle" (p. 507). If ABMs are as heuristically powerful as Goldstone and Wilensky suggest, this opens up the possibility that AC in the context of ABMs might not require the same degree of explicit instruction for learner's understanding of isomorphisms between distinct systems to be consolidated.

Summary

This chapter has laid the theoretical and conceptual groundwork for the present study by reviewing key reasons why epidemiology as a field of study poses a challenge to students and by exploring some of the most promising ways—most notably Agent Based Models (ABMs) as a teaching tool and Productive Failure (PF) as an instructional/learning approach—that this challenge might be met. The first section of the chapter surveyed ways in which the multi-level nature of epidemiology analysis (both at the level of individual agents and at the group levels of whole populations and sub-populations) along with the closely related complex systems found in epidemiology (which represent emergent properties of groups that arise from interactions at the individual level) create conceptual challenges for learners. The use of computer modelling to simulate complex system phenomena such as epidemics was then considered, with a survey of the three main types of models (System Dynamics models, Discrete Event Simulation models, and ABMs). The last of these was shown to have certain advantages over the other two—such as the ability to visualise both the individual and aggregate levels of resolution simultaneously—making ABMs a promising choice for simulating epidemics in ways that may lead to better learning outcomes for students.

The final section of the chapter included a review of the literature on PF, contrasting it to its more traditional counterpart, Direct Instruction (DI). As a learner-centred approach that allows learners to explore, fail, and use their schema abstraction before receiving guided instruction, PF was found to be well positioned to help students gain an explanatory grasp of the complex systems concepts embodying much of epidemiology and transfer that knowledge not only within previously instructed domains but to far across domains as well. Two other instructional approaches which resemble PF, Problem-based Learning (PBL) and Analogical Comparison (AC), were also found to have features that may contribute to the effectiveness of a PF approach as well. Notably, work on the interface between ABMs and AC highlights the unique heuristic power that ABMs have for fostering generalisations about deeper structural similarities across complex systems, potentially even with minimal explicit prompting or instruction (Goldstone & Wilensky, 2008; Jacobson et al., 2020).

As discussed earlier, prior studies have demonstrated the advantages of a PF instructional design across a wide range of educational fields (especially pre-university level STEM courses), and a number of studies exist which show the usefulness of integrating computer modelling into epidemiology instruction (e.g., Bai et al, 2020; Currie et al., 2020; Narassima et al., 2020). However, the intersection of PF (and similar approaches such as AC) as an instructional design model with computer simulations (ABMs, in particular) as a compatible instructional tool has received far less attention (the most notable instances being the work of Jacobson and colleagues), and investigation of this intersection in the field of epidemiology is non-existent. The present study aims to fill this gap. The next chapter presents a research methodology centred around an ABM-based, PF design curriculum employed to learn epidemics and associated complex systems concepts.

Chapter Three: Methodology

Overview

This chapter presents the details of the methodology used in the present research, a quasi-experimental study of 35 undergraduate medical students divided into two groups, each of which received three two-hour online sessions adhering to either—in the case of the experimental group—a Productive Failure (PF) format or—in the case of the control group—a Direct Instruction (DI) format. Both groups were presented with identical challenge problems modelled using Agent-Based Models (ABMs) over the course of the sessions, with related instruction presented either following or preceding the problems (PF or DI group, respectively). Pre- and post-tests assessed the students' declarative and explanatory knowledge of epidemics concepts and of complex systems concepts in the context of epidemiology. Subsequent online focus group interviews were also conducted to investigate students' overall learning experience.

Research Questions

As presented in Chapter One, the overarching research question motivating this study is whether a Productive Failure (PF) or Direct Instruction (DI) pedagogical design—each incorporating Agent-Based Model (ABM) challenge problems—leads to better learning outcomes, both for developing knowledge of complex systems concepts in epidemiology and transferring that knowledge to solve new problems. This broad question was broken down into several specific research questions (RQs) that will guide the research methodology presented in the present chapter.

The first RQ (repeated here from Chapter One) is as follows:

RQ1: Does the PF condition lead to superior learning outcomes in *declarative* knowledge of epidemics, *declarative* knowledge of complex systems concepts in epidemiology, and *explanatory* knowledge of complex systems concepts in epidemiology, as compared to the DI condition?

Given past research findings outside the field of epidemiology education indicating that PF and DI approaches are more or less equally effective at supporting the development of declarative knowledge (Jacobson et al., 2017), a nondirectional hypothesis seems most appropriate in regard to the first and second learning content components of RQ1; that is, it is expected that the PF-treatment group will not exhibit any advantage over the DI-control group on declarative knowledge of epidemics or of complex systems in epidemiology. However, in regards to the third such component—explanatory knowledge of complex systems concept in epidemiology—a directional hypothesis that the PF approach will lead to greater gains is appropriate.

The second RQ focuses on the question of knowledge transfer:

RQ2: Does the PF condition lead to superior learning outcomes in the ability to transfer knowledge of complex systems in epidemiology to new content in near within domain and in far across domains, as compared to the DI condition?

Here, again based on expectations set by prior research (Jacobson et al., 2017), directional hypotheses concerning both the near within domain and far across domain transfer components of RQ2 are warranted. That is, it is predicted that the PF-treatment group will exhibit higher gains than the DI-control group when applying knowledge of complex systems concepts to new content within both near within and far across domains.

The third RQ focuses on the session-by-session learning process:

RQ3: How does the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts affect the learning process across multiple sessions in PF vs. DI conditions?

When considering the learning process, the focus here is particularly on how an ABM-based problem-solving instructional/learning design, as presented in the context of a PF versus DI framework, may affect students' ability to apply their consolidated knowledge of complex systems concepts to real-world epidemiological scenarios and recognise deeper structural parallels and differences between the models used. Whereas the hypotheses associated with RQ1 and RQ2 can be tested based on the pretest/post-test data, RQ3 is better investigated using the students' performance on the problem-solving tasks themselves. In line with Kolchraiber et al.'s (2019) suggestion, a directional hypothesis is warranted that the PF-treatment group will exhibit better learning outcomes in their performance on the ABM-based challenge problems over the course of the three sessions as compared to the DI-control group. Unique among this study's RQs, the data associated with RQ3 will be subject both to quantitative and qualitative analysis. How this works will become clearer once the different types of challenge problems involved are explained below.

Finally, RQ4 focuses on student perceptions of their own learning experiences.

RQ4: How do students in a PF versus DI condition experience the learning of complex systems simulated via ABMs in the context of epidemiology instruction?

Being qualitative in nature, the data associated with RQ4 will require qualitative analysis to answer. Precisely how will be detailed later in this chapter.

Research Participants

Participants in the study were undergraduate students enrolled in the Faculty of Medicine at Sultan Qaboos University (SQU), Oman. (See the Data Collection section below for more information on the recruitment of participants.) Of the 35 participants who completed the study and whose data was counted (see below for information regarding participant attrition), 51.4% identified as female (n=18) and 48.6% identified as male (n=17). Overall, the average age of these participants was 20.5, with the mode participant age being

21 (n=18, 51.4%), followed by 20 (n=11, 31.4%). Most participants were in their 3rd year of study (n=27, 77.1%) and did not have English as their first language (n= 34, 97.1%). However, most participants had already been exposed to epidemiology (n= 31, 88.6%) before the start of the study. All participants selected medicine as their major. All could speak English fluently, as their language of instruction was English.

Participants were randomly assigned to one of two treatment groups, each of which worked ABM-based challenge problems, but which differed in how these problems were sequenced vis-à-vis the instructional component in each lesson. There were 20 participants in an experimental Productive Failure (PF) group, the members of which worked a selection of challenge problems each lesson before receiving instruction, and 15 participants in a control Direct Instruction (DI) group, which in each lesson received instruction before working any challenge problems (see details below on lesson design).

Regarding the 20 participants in the PF group, 10 identified as female (50%) and 10 identified as male (50%). Most members of the PF group were in their third year of study (85%), and all had Arabic as their first language, although all could speak English fluently. Finally, most of the participants had studied or been exposed to epidemiology content before (n = 20; 95%). Likewise, of the 15 participants in the DI control group, 8 identified as female (53.3%) and 7 identified as male (46.7%), with most being in either their 3rd (66.7%) or 4th year of study (26.7%). The vast majority did not have English as their first language (93%), and most had been exposed to epidemiology before (80%).

Table 1 provides an overview of the characteristics of the participant population.

Table 1

Characteristics of Participants in the Productive Failure and Direct Instruction Groups

	Total Sample (n=35)	PF Group (n=20)	DI Group (n=15)
Average Age	20.5	20.35	20.73
SD	.78	.88	.59
Age Distribution			
22	2 (5.7%)	1 (5%)	1 (6.7%)
21	18 (51.4%)	9 (45%)	9 (60%)
20	11 (31.4%)	6 (30%)	5 (33.3%)
19	4 (11.4%)	4 (20%)	-
Sex			
Male	17 (48.6%)	10 (50%)	7 (53.3%)
Female	18 (51.4%)	10 (50%)	8 (46.7%)
Year of Study			
Year 2	3 (8.6%)	2 (10%)	1 (6.7%)
Year 3	27 (77.1%)	17 (85%)	10 (66.7%)
Year 4	5 (14.3%)	1 (5%)	4 (26.7%)
Prior Exposure to Epidemiology			
Yes	31 (88.6%)	19 (95%)	12 (80%)
No	4 (11.4%)	1 (5%)	3 (20%)

It is important to note that although the final sample was 35 participants, initially 95 participants were recruited based on the number needed according to Hedge's g effect size (Sinha & Kapur, 2021; Brydges, 2019). That is, using Hedge’s g effect size from previous studies at 0.58 effect size, with α=.05, and 80% power of the study, the sample required 96 participants (i.e., 48 participants in each group). However, the data from 60 of the 95 recruited participants had to later be removed from the sample, 39 because the participants in question did not complete the entire sequence of lessons and assessments, and 21 because the participants withdrew themselves before the end of the study due to their inability to balance participation in the study with their summer courses. A post-hoc power analysis was conducted to evaluate the achieved power with the reduced sample size of 35. The analysis yielded a power value of 0.497, which indicates that the study was underpowered, as it did not meet the conventional threshold of 0.80. This achieved power reflects a lower probability of detecting a true effect, which may limit the ability to draw definitive conclusions from the data.

Research Design and Procedure

The research method was adapted from a study conducted by Jacobson and his colleagues (2017). The present study, which was conducted online with undergraduate medical students from Sultan Qaboos University (SQU), Oman, employed a quasi-experimental design with two treatment conditions: experimental PF and control DI. In addition to an online pretest and post-test, the present study’s core component consisted of three online sessions in which students explored ABMs relevant to epidemic concepts and complex systems in epidemiology during each session. AnyLogic Simulation Software was used to develop these interactive and manipulable models for learning the targeted concepts. Both groups of students (PF and DI) solved a set of identical challenge problems each session, the only difference in session design between the groups being that the PF group watched the instructional content video *after* working the second ABM challenge problem each session whereas the DI group viewed the instructional content *before* working *any* of the challenge problems.

Recent studies comparing the PF and DI approaches across multiple educational domains have found substantial evidence of the effectiveness of failure-based learning (Mazziotti et al., 2015). However, selecting an appropriate control condition remains a challenge and is an important methodological consideration that affects the validity and practical implications of research findings. In the present study, the choice of DI as a control condition is justified by empirical precedents, theoretical comparisons, and practical considerations related to the educational context of the current study.

First, use of DI for the control group in the present study allows for its findings to be more directly compared to that of a significant body of prior PF research, where DI has been the conventional choice for the control condition (Mazziotti et al., 2019; Wu, 2024a, 2024b). Moreover, this methodology explicitly acknowledges the fact that by ordering problem solving and failure before formal instruction, the PF approach represents a deliberate departure from the conventional educational sequencing (Kapur, 2014).

At a theoretical level, PF and DI methodologies have both been shown to be effective and lead to sophisticated learning outcomes (Wu, 2024a) but via different cognitive processing mechanisms. PF is typically designed as a failure-driven learning approach, in which students attempt problem-solving before instruction, whereas DI is traditionally structured to present procedures and concepts before students engage in problem-solving tasks (Kapur, 2014; Nachtigall et al., 2020). Importantly, this difference between DI and PF can be operationalised with one straightforward change of where in the order of treatment the instruction component falls, allowing the two conditions to stand in clear theoretical contrast with minimum risk of extraneous variables clouding interpretation of the results (Wu, 2024b).

Selection of DI as the control condition is also justified by the characteristics and requirements of medical education, specifically in the case of complex domains like epidemiology. PF's effectiveness varies across educational domains, where in some studies no or limited benefits have been demonstrated in non-STEM contexts (Nachtigall et al., 2020). This suggests that the choice of control condition must be aligned with the learning objectives and educational context. In epidemiology and when complex systems are being taught, the traditional instruction-centred approach is dominant, which places DI among the most relevant conditions for comparison (Lee & Lee, 2024). When dealing with complex conceptual knowledge, previous research has shown that the control condition's choice significantly influences the results' interpretability (Cao & Zhang, 2024). In medical education, using the DI control condition enables direct comparison with the most common instructional method, making it easier to determine whether the relatively greater time investment and additional complexity of PF implementation is justified.

While the use of more active control conditions might seem more sophisticated methodologically, there are limitations regarding the appropriateness and feasibility of these approaches in PF research. Alternative control conditions such as problem-positing followed by analogical comparison or instruction would introduce additional confounding factors, potentially obscuring PF's specific effects (Steinhorst et al., 2023). Moreover, there are cognitive load implications to such alternatives. Whereas PF can decrease the cognitive load on learners (Niu et al., 2021), more active control conditions may lead to competing cognitive demands, thereby masking the benefits of failure-based learning, especially in complex domains such as epidemiology.

In sum, previous PF research has provided extensive support for DI's appropriateness as a control condition, consistently showing PF vs. DI yields meaningful and interpretable results (Kapur, 2014; Suriyaarachchi, 2025). Moreover, research supports the theoretical validity of the PF versus DI comparison (Kapur, 2008). Finally, it should be noted that research on DI as a control condition has offered valuable insights into the implementation requirements and boundary conditions for effective PF interventions (Nachtigall et al., 2020).

Table 2 presents the components of each session for both groups. The experimental group followed a structured PF framework based on Kapur's (2008, 2014) design principles, which—because the instructional content was not introduced until after students had grappled with an initial set of two ABM-based challenge problems—resulted in two main phases for each session: (1) Generation and Exploration and (2) Consolidation and Knowledge Assembly.

In the first of these phases, students in the PF group worked independently for 30 minutes on challenge problem 1 using an ABM centred around basic epidemics concepts, followed by another 30 minutes on challenge problem 2 using an ABM centred around complex systems concepts. This phase encouraged students to engage in productive struggle, the hallmark of the PF approach, by attempting solutions to ill-structured problems without preparatory instruction. The goal was to activate prior knowledge, stimulate curiosity, and reveal gaps in understanding that would be addressed later. While correctness was not expected at this stage, the experience—in line with the rationale for a PF approach—primed students for deeper learning to occur during the next phase.

In the Consolidation and Knowledge Assembly Phase, the students received explicit instruction to consolidate and organise their understanding. Guided content videos recorded by an epidemiologist introduced and clarified key concepts of complex systems and epidemics related to the two ABMs with which the students had just worked, helping students develop a clearer sense of their preceding problem-solving attempts. Viewing of the instructional video was followed in each session by a final, third 30-minute challenge problem, which required students to complete two tasks: (a) first, to draw upon their consolidated knowledge (i.e., from the two ABM-based problems and the guided learning video) in order to solve a real-world problem based on the epidemics and complex systems concepts targeted in that session, and (b) then identify similarities and differences between the two ABMs, a task rooted in analogical comparison (AC) research (Gentner et al., 2003) but with Goldstone & Wilensky's (2008) proposal in mind that the rich nature of ABMs can itself be enough to guide learners' attention toward deeper structural features shared by distinct systems across domains, reducing the need for explicit prompts drawing attention to the shared structures involved. The second, compare and contrast task was included as a means of assessing whether students were, indeed, able to grasp the deeper structural features shared by the two ABMs primarily from the richness of the ABMs themselves. To that end, the researcher ensured that each pair of ABMs used in a particular session shared crucial structural parallels that students might pick up on as they worked through the two problems. (See the section "Topics and Activities Session by Session" below for a more detailed comparison of the structural features shared by each pair of ABMs; see each session's "Activities in Detail" section below for sample questions from the associated challenge problems).

Thus, the application-oriented challenge problem 3 required students to integrate and transfer their understanding, offering an opportunity to reinforce key concepts while testing students' ability to connect ideas across models. By embedding such application within the consolidation phase, the PF framework in this study aligns with Kapur's (2008, 2014) emphasis on moving from knowledge construction to meaningful application.

In the control group, the same materials and exercises were used but in a different sequence, one that reflected a traditional Direct Instruction (DI) approach. That is, during any given session, the students in the DI group first watched the same 15-minute instructional video introducing that lesson's relevant complex systems and epidemics concepts as that viewed by students in the PF group for the corresponding session. Students in the DI condition then worked the same three ABM-based challenge problems as those worked by the experimental group for the corresponding session. Thus, the only difference between the lesson designs of the two groups was the sequencing of the instructional component vis-à-vis the challenge problems. Note that because the DI group received instruction at the onset of each session, it did not experience an initial phase of productive struggle and iterative exploration in the same sense as included in the PF approach.

Table 2

Content Session Components

PF Experimental Condition

	Generation and Exploration		*Consolidation and Knowledge Assembly*	
Classroom Activities	Challenge Problem 1 Epidemics ABM	Challenge Problem 2 Complex Systems ABM	Guided Learning Video	Challenge Problem 3 Application and comparison of Epidemics and Complex Systems models
Duration	30 minutes	30 minutes	15 minutes	30 minutes

DI Control Condition

	Instruction	*Practice*		
Classroom Activities	Guided Learning Video	Challenge Problem 1 Epidemics ABM	Challenge Problem 2 Complex Systems ABM	Challenge Problem 3 Application and comparison of Epidemics and Complex Systems models
Duration	15 minutes	30 minutes	30 minutes	30 minutes

Data Collection

Both quantitative and qualitative data were collected from the students who participated in this study. The project was approved by the University of Sydney Human Research Ethics Committee, with no conflict of interest in the work being reported. Ethics approval was first received from the University of Sydney in March 2023 (Project no. 2022/583), then an ethics approval from Sultan Qaboos University, where the research was conducted, was received in June 2023. The letters of approval can be found in Appendix A. Students were emailed the participant information sheets and participant consent forms to obtain their permission before conducting the study. These forms were signed and emailed back to the researcher (see Appendices B and C).

The total numbers of SQU 3rd- and 4th-year undergraduate medical students at the time of the study was 140. To recruit participants for the study, the researcher attended one of their core classes at the invitation of the instructor and presented the research idea and parameters of involvement in the study. The students were provided with a barcode to fill in a survey that requested their consent to voluntarily participate in the study (see Appendices B and C). The survey form was also used to collect participants' email addresses for recruitment purposes and further communication.

The proposed duration of the study was two weeks; however, due to unexpected delays on the part of some participants to complete lessons in a timely manner, the actual duration of the data gathering portion of the study extended to just under three months (for some students) between June and September of 2023, with substantial attrition along the way. As noted earlier, initially 95 students agreed to participate, 48 of whom were randomly assigned to the experimental group and 47 to the control group. However, 60 of these participants failed to complete the study (most often, the third lesson and the post-test remained incomplete), leaving a final sample size of 35 (20 in the experimental group and 15 in the control group).

This substantial drop in sample size can be attributed to a number of converging factors. First, it is important to note that the research had to be carried out online due to several barriers to an in-person format. Due to the COVID-19 pandemic, many courses had been transitioned to online formats and remained that way for several years. This shift presented challenges, particularly for teaching complex systems concepts and the use of agent-based models, as the available online instructors typically lacked sufficient knowledge in these areas. Additionally, the course related to this study already covered a wide range of topics, leading to resistance from students who were unwilling to take on additional content during the regular semester. As a result, students suggested conducting the study during their summer break. However, scheduling synchronous classes during the summer proved difficult since students lived in various cities far from the university. These factors conspired to make an online format for the research the only feasible option.

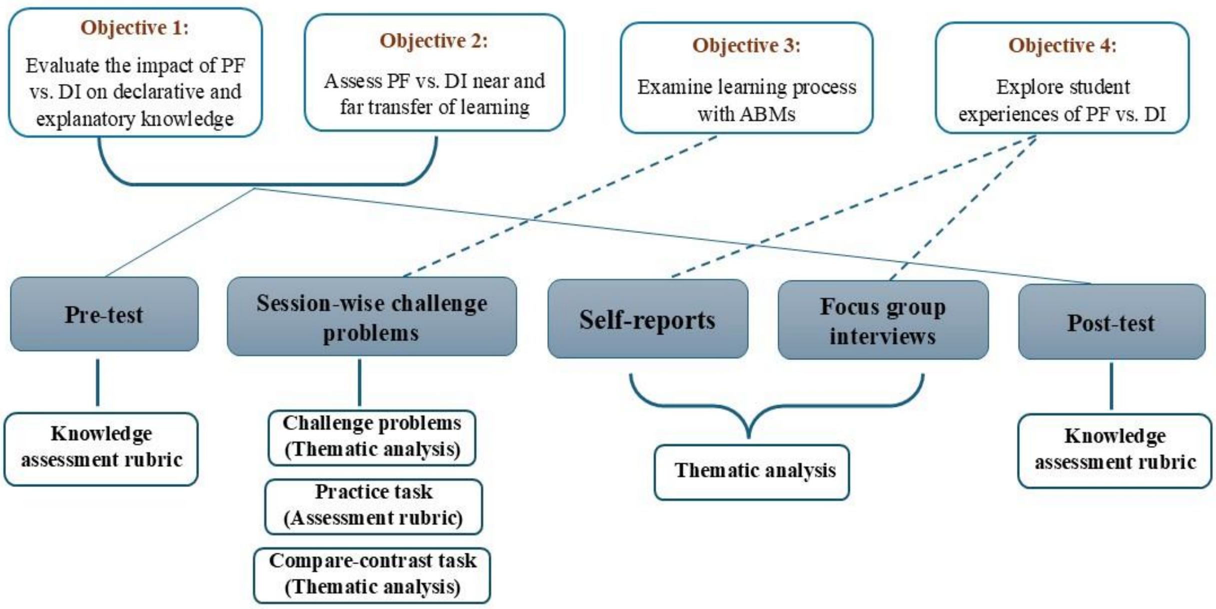
Consequently, the researcher chose to conduct the study through the online Moodle platform asynchronously during students' summer holidays. This, in turn, led to factors that seem to have contributed to the high participant attrition rate. Maintaining communication with students to complete the study materials proved difficult, as most of the participants were simultaneously pursuing summer courses or internships, hindering their progress through the study activities. Some students indicated that they had to withdraw due to such competing commitments or to health issues, whereas other students simply stopped participating. The researcher attempted to recruit additional students from the 2nd year cohort of undergraduate medical students via emails, given that these would have also already received instruction in the basics of epidemiology. However, very few such students responded to the email, and it was difficult to establish contact by other means while students were on summer vacation.

It has been observed by other researchers that relatively time-consuming interventions may result in higher dropout rates of participants. For example, in a study of exercise interventions, Moshe et al. (2022) found that 66% of their participants dropped out of the study mainly due to time-related reasons, with 40% specifically identifying limited time as the main reason for their dropping out. Previous research has also found in education settings that time constraints and demanding schedules, such as those observed in the present study, can adversely influence student retention and completion rates (Rahmani et al., 2024). Moreover, the longer the study duration, the higher the dropout rate tends to be (Gustavson et al., 2012; Rahmani et al., 2024). As noted earlier, the current study's duration unexpectedly increased from two weeks to several months due to delays on the part of participants. Indeed, one of the tasks in the study (the far across domain transfer task) was delayed eight months due to the Moodle coordinator not adding this problem to the platform and the researcher's lack of access to the platform to design and administer the course herself on Moodle. Generally speaking, as a study's duration lengthens over time, participants potentially face a greater number of constraints on their participation such as personal responsibilities, scheduling conflicts, and work commitments (Kalet et al., 2013). Moreover, sustaining motivation among participants over a longer period is challenging (Kalet et al.).

The data collected from the 35 students who completed the study came from a variety of instruments, including (a) a background entry questionnaire and instruments designed to collect both quantitative and qualitative data. Quantitative data were gathered using (b) a pretest administered before the first session, (c) a post-test administered following the third session, and (d) participants' written solutions to the ABM-based challenge problem 3 incorporated into each session. Qualitative data came from (e) the participants' written answers to the challenge problems 1 and 2 incorporated into each session, (f) the compare-contrast task presented as part of challenge problem 3 in each session, (g) participants' written self-reports at the end of each session, and (h) optional focus group interviews conducted at the conclusion of the study (see Figure 1). The data for the background questionnaire, pretest and post-test, daily challenge problems, and self-reports were all collected online via the Moodle platform and extracted as Excel spreadsheets. The focus group data were audio recorded and later transcribed with the help of the VOMO voice-to-text mobile application to then be analysed. The details of each of these data collection instruments are provided in their corresponding sections below.

Figure 1

Research Objectives and Data Collection Methods



Regarding the focus groups, all students who initially agreed to participate were emailed and asked about their willingness to join focus group interviews. Students who agreed to participate were assigned to focus groups based on their membership in the experimental versus control group and based on gender (the focus groups were separated based on gender in light of cultural factors and at the request of female students so that they could more freely discuss and express their opinions). This procedure resulted in the formation of four focus groups, with interviews to be administered online via Zoom. Two groups of students were formed from members of the experimental group who agreed to participate in the interviews (5 male; 3 female), and two groups of students were likewise formed from members of the control group (3 males; 4 females). However, the meeting of the male focus group from the control group participants was cancelled, as two of the three members of that focus group failed to appear at the agreed upon time. Nonetheless, written responses from the members of this focus group were requested and received, and these were used in the subsequent analysis. The time duration of each focus group discussion that met on Zoom was 45 minutes. More details are provided below in the focus group interview section.

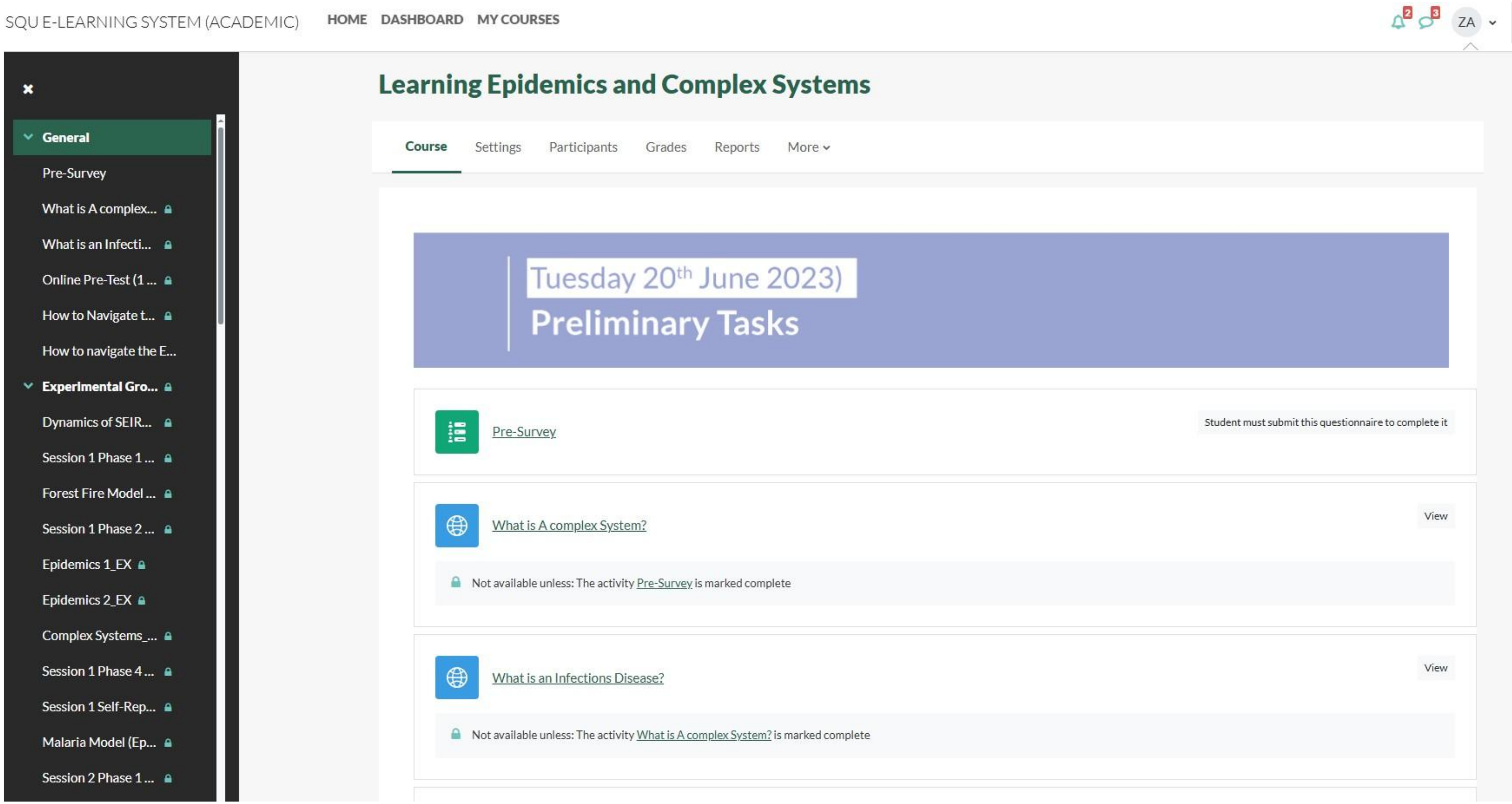
Instructional Environment for the Three Content Sessions

Course Shell on Moodle Platform

An asynchronous online course shell (Figure 2) was designed on the Moodle platform belonging to Sultan Qaboos University (SQU), where the participants study and have accounts for access. A Moodle coordinator from SQU was responsible for organising the course, tracking students' progress, notifying them as needed, and finally sharing the data sets with the researcher. The course included a tutorial video created by the researcher and tailored for each group (PF and DI) explaining the overall structure of the courses, including the sequence of tests, sessions, and tasks, and how to access the ABMs that were used to solve the challenge problems, with links to the ABMs provided. (The sequence was different between the groups, so separate videos were required for this purpose.) The course also included the instructional content videos (one for each session, the same for both groups), general guidelines for using the ABMs to solve the associated challenge problems, and detailed instructions for each problem. (See below and Appendix D for more details). The time limit for each task was displayed on screen when students solved problems, and a feature was activated that restricted students from browsing other web pages during the task completion. Students were notified via their emails frequently to remind them of tasks still needing completion.

Figure 2

A Screenshot of the Study Course Designed on the SQU Moodle Platform



ABMs

Six ABMs were designed and developed in AnyLogic (https://www.anylogic.com) by the researcher to be used for solving challenge problems related to epidemics and complex systems content. AnyLogic is "the leading simulation modelling software for business applications" and offers a more engaging interface than the NetLogo program often used in educational contexts, although unlike NetLogo, AnyLogic does require some knowledge of Java for the creator to develop "more robust, extendable, and reliable models" (Vosloo & Lemos, 2021). Using AnyLogic, the SEIRS Dynamics, COVID-19, and Malaria models used in this study were created for learning epidemics concepts (one model for each of the three content sessions), while the Wolf-Sheep Predation, Marketing, and Forest Fire Models were designed for learning complex systems concepts (again, one model for each session). Research participants were provided with online guidelines illustrating the general procedure for using the ABMs to solve the challenge problems (Figure 3) along with detailed instructions for each problem (Figure 4).

Figure 3

General Guidelines for How to Use ABMs to Solve Challenge Problems

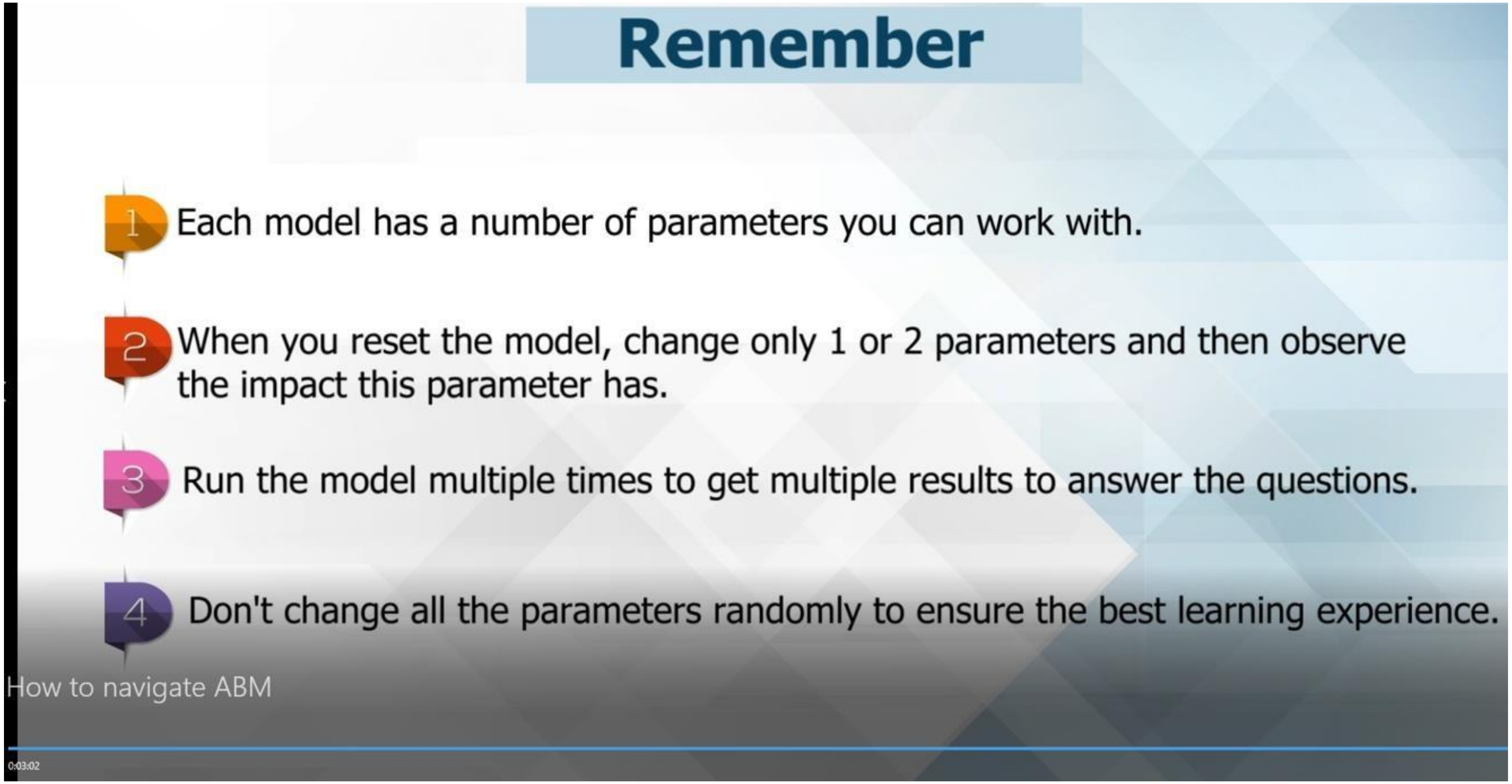
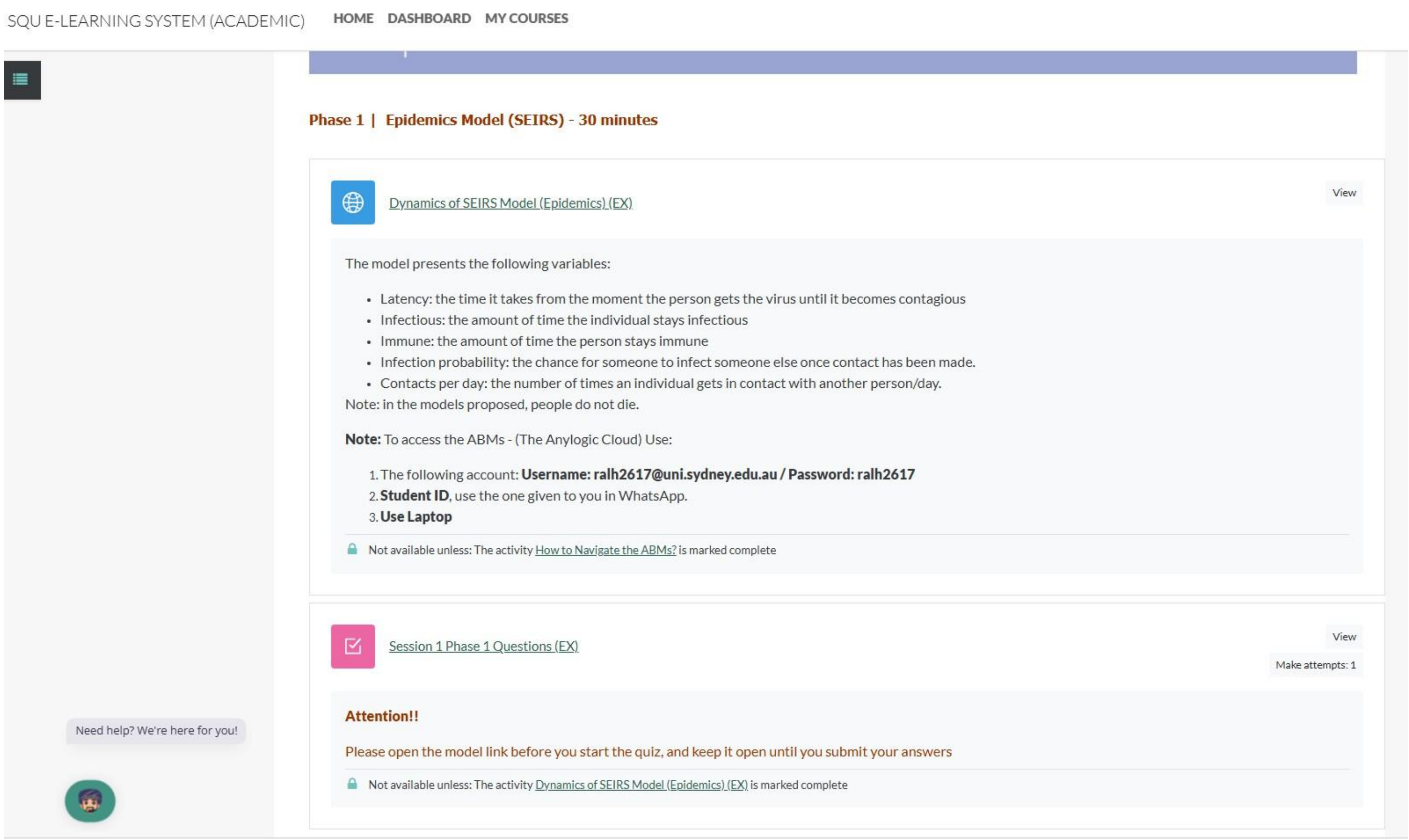


Figure 4

A Screenshot Showing How Instructions Were Provided for Each Task



Instructional Videos

Each of the three sessions in this study featured a guided learning video that included an explanation of the epidemics and complex system concepts for that lesson. Scripts for these videos were developed by the researcher and reviewed by Michael Jacobson, who acted as an academic supervisor (See Appendix E). The video was then recorded by Dr. Gina Arena, a senior lecturer at the University of Western Australia Medical School (Figure 5).

Figure 5

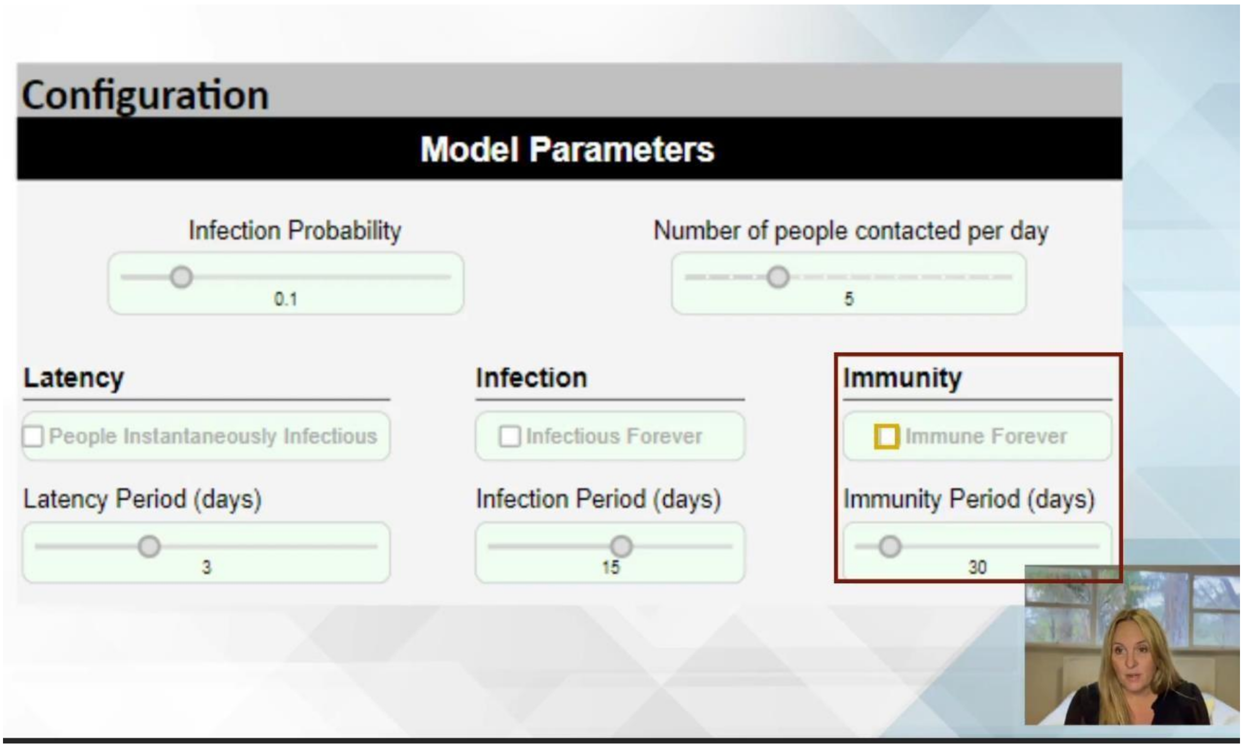
A Screenshot of the Guided Videos Recorded by Dr. Gina Arena

It should be noted that, ideally, students in a PF learning environment would be able to discuss with their instructors the solutions they generate in the process of problem solving, as this is considered an important design principle of the PF approach (Loibl et al., 2017; Loibl & Rummel, 2014a). However, as mentioned earlier, there were a number of mitigating factors that prevented the present research from being carried out in an in-person format.

Given that there were no face-to-face classes involved in the present study in which to initiate student-teacher discussions, some of the results expected from the models were instead provided in the videos themselves (Figure 6).

Figure 6

An Example of the Expected Answers Discussed in the Videos



Topics and Activities Session by Session

The pretest and post-test were both administered online, with the subjects also completing the background questionnaire and watching the tutorial video at the time of the pretest. Between the pretest and post-test were the three content sessions in which students engaged in ABM-based problem solving and viewed instructional videos. Table 3 summarises the conceptual content covered in each of the three content sessions along with the names of the relevant ABMs. Each session, students were asked to solve three challenge problems using two relevant ABMs in each session. They used an epidemics model to solve challenge problem 1 and a complex systems model for challenge problem 2. Students used both the epidemics and complex systems models explored in problems 1 and 2 to work on challenge problem 3, which required them to apply the concepts presented in the previous two challenge problems (and discussed in the associated instructional video) to solve a related real-world problem and also identify similarities and differences between the two models. As noted earlier, these activities were held in common between the PF and DI conditions, the only difference between the two groups being the order in which the instructional video was presented relative to the challenge problems.

It is important to note that in every session, conceptual content on both complex systems and epidemics was presented. This was a deliberate design choice, grounded in these domains' inherent interrelatedness. In complex systems, epidemics are considered as quintessential examples, being characterised by emergent phenomena, non-linear behaviours, and dynamic interactions; this makes it pedagogically meaningful to consider their shared significant characteristics, including system-level outcomes, agent interactions, and feedback loops (Uskola & Puig, 2023; Yoon et al., 2023). The curriculum used in this study allows students to identify and evaluate cross-cutting concepts (by engaging them concurrently with both subjects), such as adaptive behaviour, system structure, and causality that are crucial to learning real-world medical challenges (Yoon et al., 2023; Uskola & Puig, 2023).

Research shows that when underlying principles overlap, understanding one complex domain facilitates transfer to another, specifically when learners are encouraged to draw connections between challenge problems that are related (Lai et al., 2017; Uskola & Puig, 2023). The challenge problems' order was intended to support this transfer, that is, one domain's initial exposure (i.e., epidemics) offered students a conceptual foundation, which could then be leveraged for the subsequent domain (i.e., complex systems). This sequencing deepened students' understanding and enhanced their ability to flexibly apply knowledge (Lai et al., 2017; Guo et al., 2013). The sequencing also aligns with previous literature calls to design educational interventions so as to foster both far and near transfers, especially in ill-structured and interdisciplinary contexts (Uskola & Puig, 2023; Lai et al., 2017). The focus on declarative knowledge in both complex systems and epidemics was, therefore, directly relevant to the current research's objective of comparing PF vs. DI's efficacy. This arrangement enabled a robust analysis of whether PF better enables the acquisition of foundational knowledge and the integration and transfer of that knowledge across domains—a crucial competency in medicine, where systems thinking and adaptability are critical (Yoon et al., 2023; Sinha & Kapur, 2021).

Table 3

Conceptual Content for Both Groups by Session

Content Session	Epidemics ABMs & Concepts	Complex Systems ABMs & Concepts
1	SEIRS Model: • Dynamics of SEIRS • Latency • Prevalence • Infectivity • Mortality	Forest Fire Model: • Emergence • Systems' levels (micro and macro)
2	Malaria Model: • Human-vector transmission • Mitigation policies	Marketing Model: • Tipping points • Linear and non-linear changes
3	COVID-19 Model • Human-human transmission • Mitigation policies	Wolf-Sheep Predation Model: • Dynamic equilibrium

As Table 3 shows, Content Session 1 focused on the complex systems concept of emergence and the closely related phenomenon of interaction between agents at a micro level generating system properties at the macro level. In an epidemiological context, emergence corresponds to the dynamics of epidemics in a population, particularly as captured by the SEIRS theoretical approach (Susceptible-Exposed-Infectious-Recovered-Susceptible; Bjørnstad et al., 2021). Session 2 introduced the complex systems concepts of tipping points, linear changes, and non-linear changes. The first of these references a threshold moment when a system dramatically or rapidly changes from one stable state to another. Such shifts in state usually occur as a response to a small change or the accumulation of previous small changes. Changes in a system that are proportional to the input that caused them are described as "linear" changes, whereas less predictable, nonproportional changes are referred to as "non-linear" changes (Hardesty, 2010). Session 3 focused on the dynamic equilibrium concept, which refers to a system's ability to maintain its overall stability despite disturbances of the elements. The complex systems concepts covered in Sessions 2 and 3 are related to the transmission dynamics of two diseases (Malaria and COVID-19, respectively) in the context of epidemics. Table 4 displays the learning objectives of each day of the study.

Table 4

Learning Objectives by Session

Session	Learning Objectives
1	• Evaluate the dynamics of epidemic diseases and their application in a population through SEIRS.
2	• Explain the modes of transmission for malaria. • Evaluate the methods for preventing malaria. • Compare the preventative measures for the spread of malaria. • Evaluate the effectiveness of the policies used for reducing malaria.
3	• Explain the modes of transmission for COVID-19. • Evaluate the methods for preventing COVID-19. • Compare the preventative measures for the spread of coronavirus. • Explain the different global responses related to coronavirus (including policies).

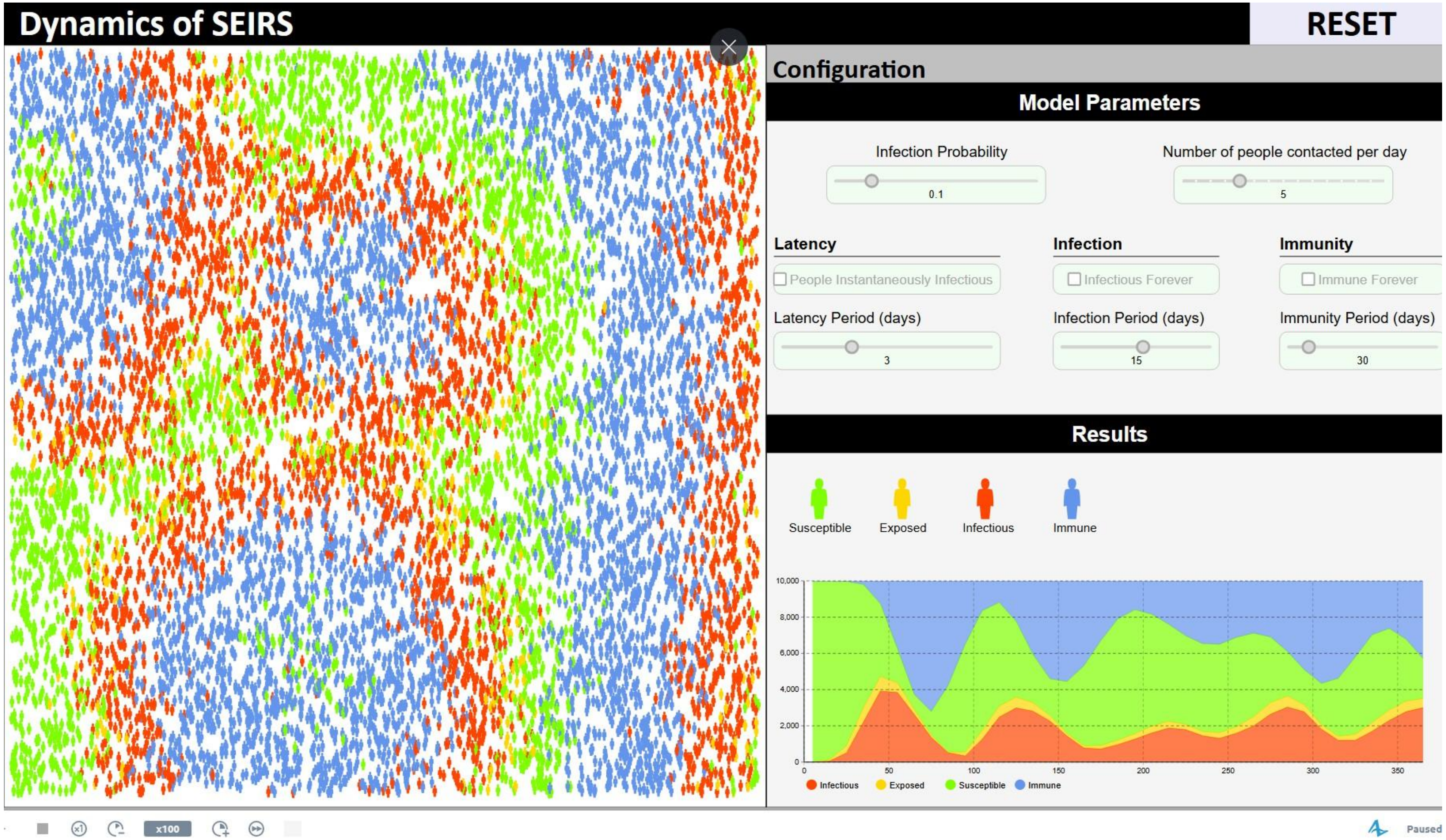
Content Session 1 Activities in Detail

As with the other content sessions below, the first content session, Session 1, began differently for the PF experimental condition and the DI control condition. The DI group's Session 1 began with the corresponding instructional video created for this session, which presented and explained the epidemics and complex systems concepts that would be practiced in the challenge problems to follow. In contrast, the PF group's session began by presenting the first of the challenge problems described below, the instructional video being held until later in the session (i.e., after the second challenge problem).

The first challenge problem given to the students in Session 1 concerned the dynamics of epidemics and how theoretical models, particularly SEIRS, can be applied to investigate the behaviour of viruses within a population. Students explored the parameters provided in the associated ABM by ticking/unticking boxes or increasing/decreasing the numbers to investigate how changes in variables could impact the prevalence of the virus within a community. Students were required to run the model multiple times and take notes, as shown in the results chart and the animation (Figure 7). Based on the SEIRS model, this session mainly focused on the following parameters: latency, infection probability, contact per person, mortality, and immunity.

Figure 7

A Screenshot of the SEIRS Model Interface



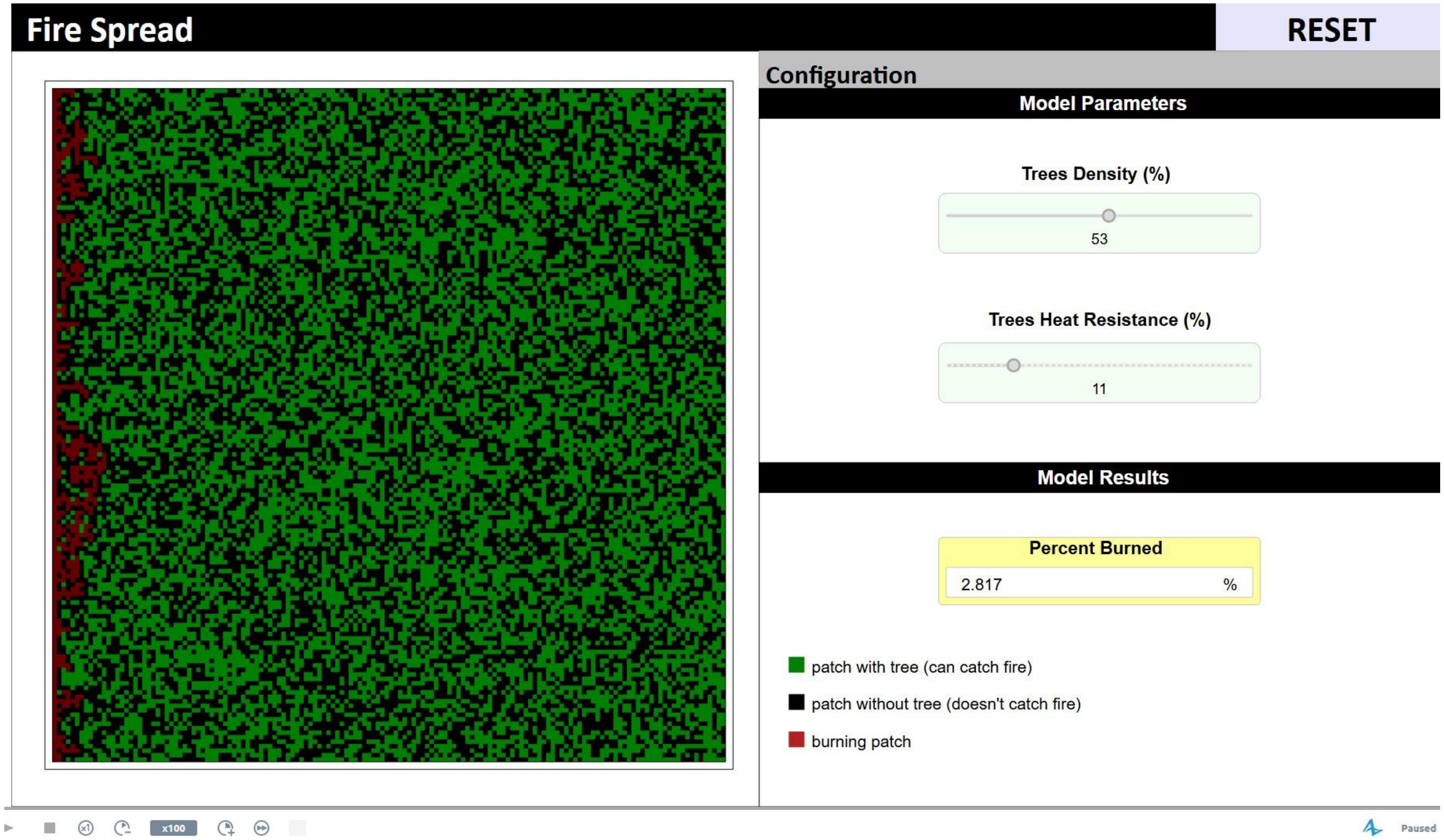
As the culmination of their exploration of the model, students were required to answer the following question:

> How is it that the population constantly gets infected and the disease continues to spread over time? Write down as many ideas as you can come up with, that might answer this question. It is not important to identify a “right” answer.

In the second challenge problem of Session 1, students were provided with a Forest Fire ABM consisting of a forest with trees of a certain density. They used this model to navigate relevant parameters and observe the changes that occurred within the forest (Figure 8). The complex systems concepts involved were emergence and systems’ levels, these being related to how the micro-level interactions of one burning tree spreading to another close to it can result in a macro-level forest fire.

Figure 8

A Screenshot of the Forest Fire Model Interface



As the culmination of their exploration of the model, students were required to answer the following question:

> The forest world is built in the simulation with pixels. Each pixel (or patch) represents either a tree, fire, or empty space. When you watch the simulation run, it looks like the fire moves, nevertheless, none of these pixels (or patches) actually move. Write as many ideas as you can come up with, that might answer why the fire never actually moves. It is not important to identify the “right” answer.

Note that at this point in Session 1 (following the second challenge problem), the PF group was shown the instructional video, which clarified the concepts covered above in the first two challenge problems and their associated ABMs.

The session ended for both the PF experimental group and the DI control group with a third challenge problem, which required students to use both of the ABMs explored in challenge problems 1 and 2 so as to compare the students’ initially generated solutions to the canonical ones that had been explained in the content video. Two open-ended questions were posed to students in challenge problem 3, the first requiring them to apply the concepts learned from the preceding portions of the lesson to a novel real-world scenario, and the second requiring them to consider the similarities and differences between the two simulation models from that lesson.

1. You are a scientist who works in the public health department of a Middle Eastern country. The Minister of Health has asked you to advise on the dangers associated with a highly infectious disease that has recently been identified, with unknown characteristics. Provide the Minister a summary of your assessment on how the disease is spread amongst the population depending on the disease characteristics. Please present your summary in general terms so that the public would understand.

2. Write down what you think are the main similarities and differences between the two models you have looked at.

Content Session 2 Activities in Detail

As with the preceding session, the second content session began either with the instructional video created for this session (in the DI condition) or with the first of this session's challenge problems (in the PF condition). The first challenge problem was based on the Malaria ABM (Figure 9), in which epidemics concepts were introduced that went beyond the SEIRS model used in the previous session—for example, incubation period, hospital and ICU capacity, and preventive policies such as use of repellent, adequate clothing, and mosquito nets. As the hypothetical malaria epidemic worsens, the ABM animation keeps track of whether individuals are infected or not, where they are located at any given time (at home, outside, hospitalised or in ICU), and whether the hospital or ICU becomes full. If a person requires hospitalisation (or ICU) and there is no space available, there is a certain period after which the person may die. Other variables are taken into account as well; for example, a person who is inside a house is at less risk of being exposed to mosquitoes. Students were encouraged to run the model many times with different changes in the parameters to examine the transmission patterns of Malaria and the ways mitigation policies could help reduce the prevalence of the disease.

Figure 9

A Screenshot of the Malaria Model Interface

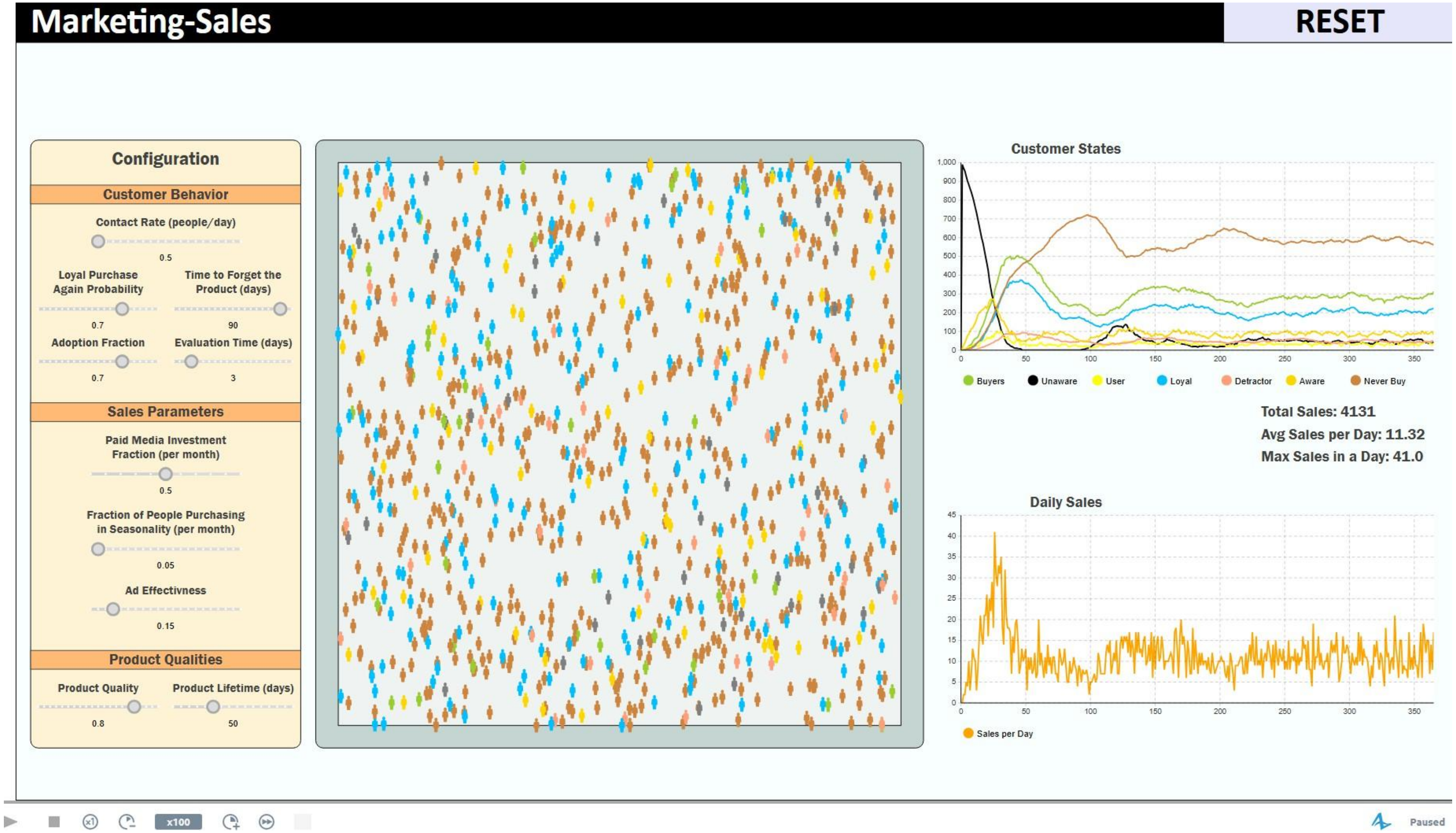
Students then answered an open-ended question related to the scenario:

> Explain how the disease spreads over time and the population's behaviour towards this. Write as many ideas as you can come up with, that might answer this question. It is not important to identify the "right" answer.

The second challenge problem of Session 2 used the Marketing ABM to explore the complex systems concept tipping point in the context of marketing campaigns such as advertisements for medical equipment. In the ABM, an individual becomes aware of the product in a certain way and then will either buy or not buy the product. On the results side of the window (Figure 10), there are two-time plots: one that shows changes in the state of customers over time and the other the number of sales per day over time. Based on the changes seen in these plots, the students predict the tipping points. The main parameters in this model are contact rate, adoption fraction, product quality, and product lifetime.

Figure 10

A Screenshot of the Marketing Model Interface



Again, the students were required to answer an open-ended question related to the model:

> Explain the people choices over time and the factors that affect the population's behaviour. Write as many ideas as you can come up with, that might answer this question. It is not important to identify the “right” answer.

As in the previous session, following the second challenge problem the PF group was presented with the instructional video for this session. (The DI group had already viewed this video at the beginning of the session.)

The final activity for both groups was a third challenge problem, which asked students to address a real-life scenario and compare/contrast the epidemics and complex systems concepts from the two ABMs used in the session. These two open-ended questions were as follows:

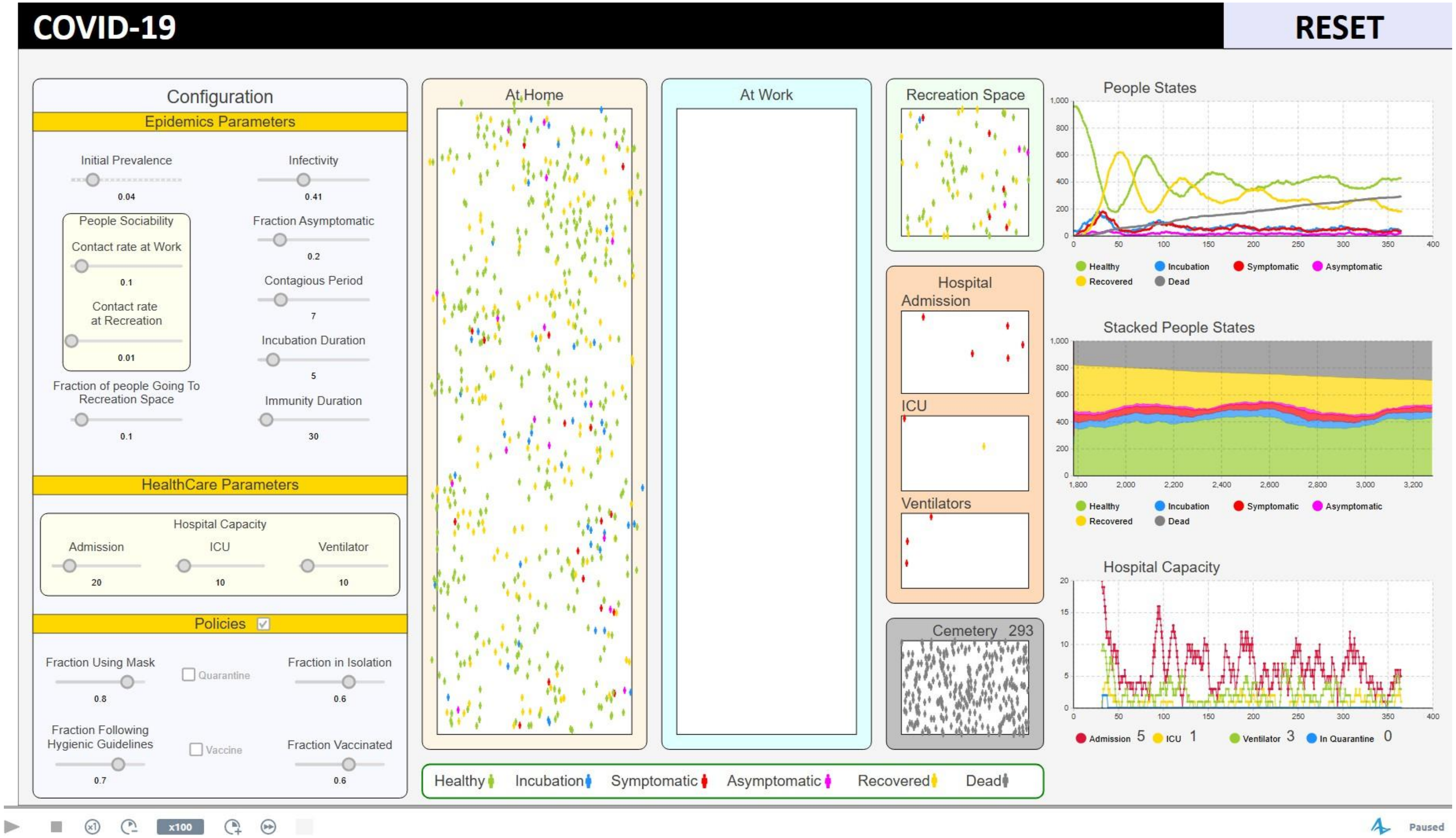
1. You are a researcher working at the local university and you are asked for recommendation policies to implement in response to the Malaria situation based on your current research findings. Think about the recommendations you can make and their implication in the way the population will react compared to the results in reducing the spread of this disease. Explain your reasoning for a non-technical audience, using epidemics and complex systems concepts as to why your policy recommendation will be effective in the short and/or long term.
2. Please explain how the concept of Tipping Points applies to the two models you have investigated.

Content Session 3 Activities in Detail

Like the other content sessions, Session 3 started either with the relevant instructional video (DI condition) or the first challenge problem (PF condition). The latter was based on the COVID-19 ABM, which highlighted the epidemics concepts of human-human transmission and mitigation policies (Figure 11). Given the nature of the coronavirus, this ABM required more advanced parameters and animations to deliver a holistic image of the typical behaviour of the disease spread. Three categories of parameters were needed: (a) general parameters, including infectivity, the fraction of asymptomatic people, the contagious period, incubation duration, and immunity duration; (b) sociability parameters, which focused on the contact rate at work and in the recreation space; and (c) healthcare parameters, which included hospital admission, ICU admission, and the availability of ventilators. Students were guided to initially run the model multiple times without ticking the policies box to help ensure they perceived the changes over time and took notes. They were then instructed to tick the policies box and run the model again to see how the stated preventive measures could contribute to the control of the disease within the population.

Figure 11

A Screenshot of the COVID-19 Model Interface



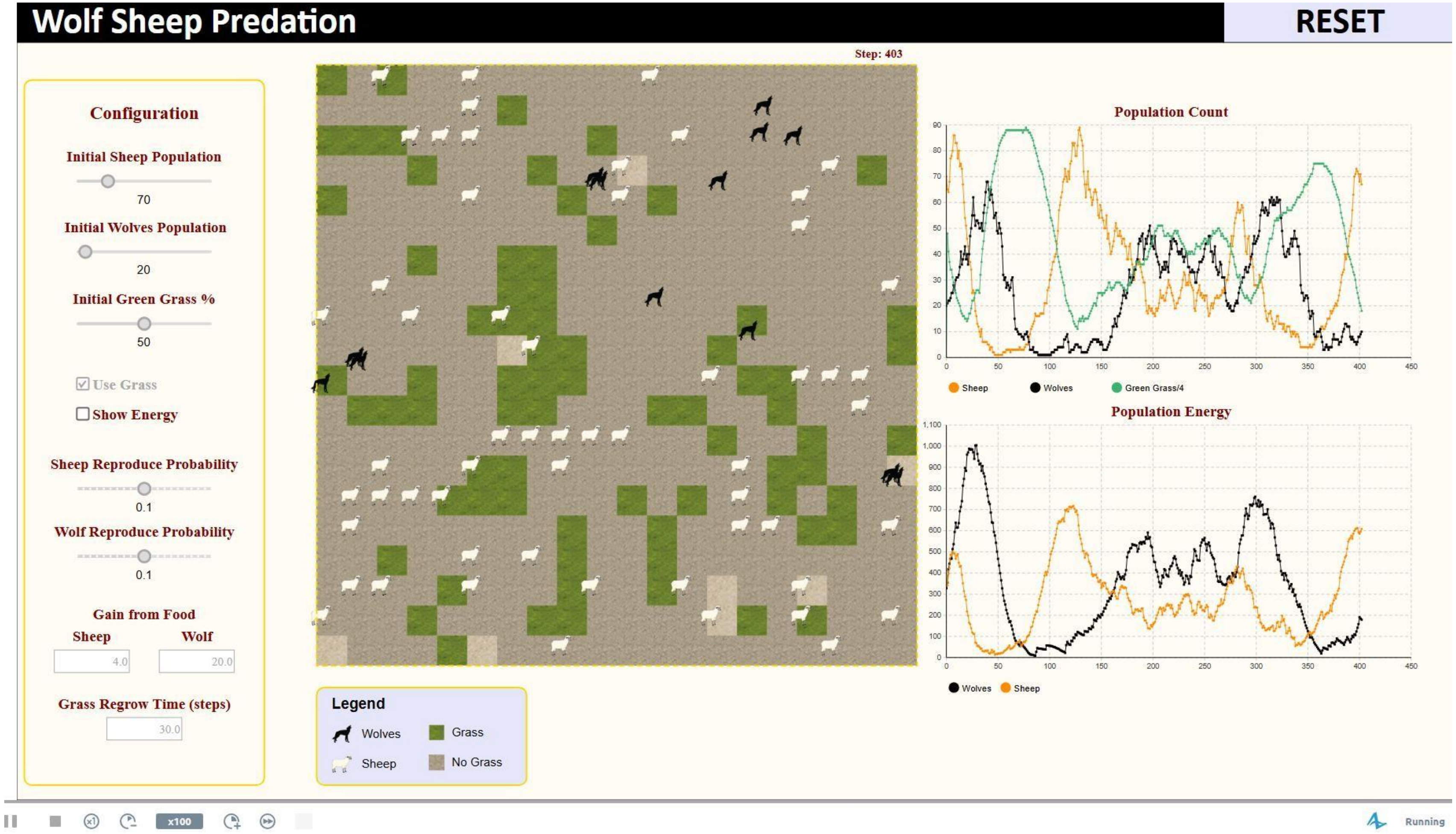
Following their exploration of the model, students answered the following open-ended question:

> Explain how the disease spreads over time and the factors that affect the populations' behaviour? Write as many ideas as you can come up with, that might answer this question. It is not important to identify the “right” answer.

The second challenge problem of this session used the Wolf-Sheep Predation ABM to explore the concept of dynamic equilibrium. This ABM demonstrated the dynamics between predators and prey in an ecosystem, how they influence each other's population sizes, and how equilibrium is established over time. The model interface included a list of parameters and two graphs that represent the changes in the number of animals and in the animals’ energy levels over time (Figure 12).

Figure 12

A Screenshot of the Wolf-Sheep Predation Model Interface



After exploring the model, students answered the following two questions:

1. How is it that the sheep do not eat all the grass or that the wolves do not eat all of the sheep? Write down as many ideas as you can come up with that might answer this question. It is not important to identify the "right" answer.
2. How would you describe the type of relationship that you observe between grass, sheep, and wolves?

Following the second challenge problem, the PF group was presented with the instructional video for this session.

Finally, both groups of students solved the third challenge problem, which required them to use what they learned from the preceding two models to answer the following questions:

1. As a Director of the Inpatient ward at the local hospital, analyse the way hospitals respond during an increase in symptomatic populations. Propose some appropriate recommendations for the healthcare professionals to reduce the mortality in the population.
2. Write down what you think are the main similarities and differences between the two models you have looked at.

Data Sources

Multiple sources of data were gathered online prior to, during, and at the conclusion of the study, including (a) written responses of participants' background information, (b) written responses of the pretest and post-test, (c) written responses to the challenge problems completed as part of the content sessions, (d) student self-reports at the end of each session, and (e) focus group interviews at the conclusion of the study. Each of these will be considered in more detail below.

Background Questionnaire

The background questionnaire was completed by potential participants online in the Moodle platform. The questionnaire asked for respondents' gender, age, year of study, and level of knowledge of epidemiology (See Appendix F).

Knowledge Assessment: Pretest and Post-test

The pretest was administered prior to Session 1 and the post-test after Session 5 using the Moodle platform to assess students' declarative and explanatory knowledge of complex systems and epidemics concepts as well as their ability to transfer acquired knowledge to new contexts. Part One of both the pretest and post-test included seven open-ended questions designed to assess (a) declarative knowledge of epidemics concepts (i.e., questions 1-4), (b) declarative knowledge of complex systems concepts in epidemiology (questions 5A-7A), and (c) explanatory knowledge of complex systems concepts in epidemiology (questions 5B-7B). Representative questions from both tests are listed in Table 5, which categorises what type of knowledge each question measured.

Part Two of the pretest and post-test required students to solve problems that demonstrated their ability to transfer what they had learned in the sessions to new problems within the domain of epidemiology (i.e., near within domain transfer). Specifically, one problem questioned students on the difference between an "epidemic" and an "outbreak" of a disease in regard to cases of the common cold versus tuberculosis in a specific timeframe and region (the "Common Cold" problem in Table 5) while the other problem questioned students about transmission of and possible preventative measures for cases of salmonellosis in a population (the "Salmonella Enterica" problem in Table 5). The post-test additionally included a problem in Part Three whose solution required far across domain transfer of knowledge to an entirely new domain outside of epidemiology, namely, mobile phone use as it relates to road injury and mortality rates in Oman (the "Road Injury and Mortality" problem in Table 5). All the transfer problems on the pretest and post-test focused on topics that were not previously included in the study, requiring students to apply (i.e., transfer) epidemics concepts that had been explored—such as prevalence, mortality, and mitigation policies—to reach appropriate solutions.

The study categorised 'common cold' as an example of near transfer based on similarity of contexts. Preventive behaviours related to the common cold (e.g., hand washing) are typically practiced and taught in environments similar to those in which preventative behaviours for epidemics are required, such as public places, schools, and homes (Okunlola, 2023). On the other hand, 'road injury and mortality' is an example of far transfer based on dissimilarity of contexts, given that the behaviours required to prevent road injuries and mortalities (e.g., observance of traffic signs) are often taught in simulated or controlled environments before being applied in real-world situations. Thus, in the case of the common cold, there is task overlap: skills and knowledge related to epidemics are applicable directly to the prevention of common colds. In contrast, in the case of road injury and mortality, there are contextual and task differences (HosseiniKhezri et al., 2025). All of the transfer questions in view were open-ended and required the students to write explanatory answers (see Appendix F for the details of these problems). To ensure the tests' face validity, a university epidemiology specialist reviewed the questions and approved their relevance to the targeted content.

Table 5

Sources, Types of Information, and Sample Questions

Source	Type	Questions
Pretest & Post-test	Declarative knowledge of epidemics concepts (Questions 1-4)	1. What are the modes of transmission of infectious organisms? Why is it important to know them? 2. What is the difference between latency and incubation period? 3. What does mortality refer to in epidemics? 4. Please describe what disease prevalence is.
	Declarative knowledge of complex systems concepts in epidemiology (Questions 5A-7A) Explanatory knowledge of complex systems concepts in epidemiology (Questions 5B-7B)	5A. What are examples of emergent properties in epidemics? 5B. Please explain. 6A. What are examples of tipping points in epidemics? 6B. Please explain. 7A. What are examples of dynamic equilibrium in epidemics? 7B. Please explain.
Pretest & Post-test	Near within domain transfer problems	1. "Common Cold" problem Key questions: (a) Discuss the disease and its ability to move through the population. (b) With the information that you have regarding this disease, what preventative measures can be put into place?

Source	**Type**	**Questions**
		2. "Salmonella Enterica" problem Key question: What are the differences between an epidemic and an outbreak of a disease in relation to the different population sectors involved in the proliferation of the disease?
Post-test	Far across domain transfer problem	3. "Road Injury and Mortality" problem Key questions: (a) Can you explain the relationship between mobile phones, road injury, and mortality? How does prevention relate to data surveillance? (b) Please describe and explain why different preventative and regulatory measures are used for different members of the population to reduce the rate of injury and mortality associated with mobile phones?

Challenge Problems

As already described in above in some detail, students completed three ABM-based challenge problems during each content session 1-3 to explore the targeted epidemics and complex systems concepts in each lesson (refer to Appendix D). In regard to data generation and analysis, the data that resulted from these exploration, consolidation, and practice phases of each session (depending on the experimental or control condition involved; refer to Table 2) allowed for a closer examination of the ABM-based learning process within and between the PF and DI groups.

Self-reports

Participants of both groups were encouraged to self-report their own experience in regard to the online sessions and similar experiences in the past. Specifically, students answered two open-ended questions after each session on the Moodle platform (See Appendix G):

- Think about how your learning experience of epidemics and complex systems concepts went in this and previous sessions. Provide your feedback on what you think works well for you and what parts could be improved.
- Think about how your usual learning experience of epidemics or any other theoretical medical courses normally go. Provide your feedback on what you think works well for you and what parts could be improved.

The purpose of asking these questions was to receive qualitative data that might support interpretation of the results of the knowledge assessment.

Focus Group Interviews

Three focus group interviews (Neuman, 2014) were administered via Zoom with a portion of the students from both treatment groups in order to understand their perspectives and attitudes toward the instructional approaches and integration of ABMs in learning complex systems and epidemics. The interview with each group lasted approximately 45 minutes. Each of these Zoom meetings started with a welcome and a brief overview of the purpose of the meeting. This was followed by presentation of the "ground rules" so that all participants understood there were no right or wrong answers and that all perspectives were welcomed. Participants were informed that their discussion would be audio recorded for analysis purposes and that their names would be kept confidential. Participants were then questioned regarding how they felt about the ABMs used during the sessions, how their experience in the sessions compared to more traditional science teaching methods, whether they found the complex systems ideas included in the sessions helpful for understanding epidemiology concepts, and whether the approach used in this study might be helpful for learning other medical and health concepts. Students were also given the opportunity to comment on any aspect of the learning activities used in the present study. (See Appendix H for a list of the questions presented in the focus group interviews.)

The questions and follow-up probing of the students' answers were intended to support a richer understanding of the learning process and the students' perspectives of it. The qualitative data from the interviews were initially reviewed and then transcribed using the voice-to-text VOMO mobile application. The researcher reviewed the scripts to ensure their accuracy.

Data Analysis and Ethics

RQ1: Hypotheses and Method of Analysis

Hypotheses for RQ1. The first research question of the present study focuses on whether the experimental treatment group (PF) and control group (DI) will exhibit differences in their quantitative gains on declarative and explanatory knowledge of epidemics and complex systems ideas. Three types of knowledge in this regard are queried on the pretest and post-test (refer to Table 5):

1. *declarative* knowledge of epidemics (pretest/post-test questions 1-4),
2. *declarative* knowledge of complex systems concepts in epidemiology (pretest/post-test question 5A-7A), and
3. *explanatory* knowledge of complex systems concepts in epidemiology (pretest/post-test questions 5B-7B).

Two hypotheses were developed in association with RQ1:

H1 (nondirectional): The PF-treatment group will not exhibit better learning outcomes than the DI-control group on *declarative* knowledge of epidemics or *declarative* knowledge of complex systems in epidemiology.

H2 (directional): The PF-treatment group will exhibit better learning outcomes than the DI-control group on *explanatory* knowledge of complex systems in epidemiology.

Choice of **d***ata* **a**nalysis *for RQ1.* To examine the differences in outcome scores on the post-test between students in the PF treatment condition and those in the DI control condition on the above three types of knowledge, the initial plan was to conduct a Multivariate Analysis of Variance (MANOVA), as the MANOVA is typically appropriate when multiple dependent variables that are theoretically related are examined simultaneously (Stockburger & Frey, 2018). In this study, the outcome variables share conceptual linkages given that they all assess various dimensions of knowledge acquisition within the context of epidemiology education. The use of MANOVA, as opposed to conducting separate t-tests or ANOVAs for each dependent variable, would provide two primary advantages. First, it would reduce the risk of Type I error inflation, which occurs when multiple tests are conducted independently on related outcomes (Pituch & Stevens, 2015). Second, MANOVA allows for the examination of group differences on a linear combination of dependent variables, enabling the detection of broader patterns that may not emerge when each variable is analysed in isolation (Pituch & Stevens).

Before conducting the MANOVA, preliminary analyses were performed to ensure that the necessary statistical assumptions for the MANOVA were met. The first assumption, independent random sampling, requires that observations are independent, with no systematic pattern in the selection of participants, and that the sample is chosen randomly (Pituch & Stevens, 2015). This assumption was satisfied, as the students were randomly assigned to either the experimental condition (PF) or the control condition (DI). The random assignment process ensured that there was no discernible pattern in how participants were allocated across the two instructional conditions.

The second assumption for conducting a MANOVA pertains to the level and measurement of the variables. That is, the independent variables must be categorical, and the dependent variables must be continuous or measured on a scale. This assumption was also satisfied, as the independent variable—representing the PF and DI conditions—is categorical, and the dependent variables were measured on a 4-point scale, as indicated by the assigned coding scheme (see "Scoring of Quantitative Data" below).

The third assumption for conducting a MANOVA is the absence of excessive correlation between the dependent variables. According to Chan et al. (2022), no correlation should exceed $r = .90$. To test this assumption, a Pearson r correlation analysis was conducted on the post-test scores, as these are the values included in the MANOVA model. The results, presented in the correlation matrix in Table 6, revealed that declarative knowledge of complex systems in epidemiology and explanatory knowledge of complex systems in epidemiology was highly correlated ($r = .96$). However, in many research contexts, high correlation between dependent variables may occur naturally, especially in the case of variables measuring related constructs. Still, previous research suggests that among dependent variables, presence of high correlations may result in challenges in interpreting MANOVA's results, especially while evaluating group differences' significance across these variables (Tonidandel & LeBreton, 2013). This suggests that MANOVA would be unsuitable for the present dataset.

Table 6

Correlation Matrix for Dependent Variables

Variable	(1)	(2)	(3)	(4)	(5)
(1) Declarative epidemics knowledge	-				
(2) Declarative complex systems in epidemiology	.56**	-			
(3) Explanatory complex systems in epidemiology	.63**	.96*	-		
(4) Near transfer	.53**	.43**	.51**	-	
(5) Far transfer	.39*	.06	.17	.43**	-

*Note: *p<.05; **p<.01*

Given this limitation, the remaining assumptions were tested to determine whether an alternative analysis could be conducted. Specifically, if the assumptions of normality and homogeneity of variance were satisfied, multiple independent samples *t*-tests could be performed. Unlike the MANOVA, *t*-tests do not require the absence of excessive correlation between the dependent variables, as each dependent variable is analysed separately. Furthermore, studies suggest that when MANOVA is not appropriate, independent samples *t*-tests serve as a suitable alternative, particularly when there are only two conditions and the other relevant assumptions are met (Huberty & Olejnik, 2006).

The next assumption tested was that of normality. This means that the data should follow a normal distribution for both the PF and DI groups. It is important to note, however, that prior research indicates the independent samples *t*-test is relatively robust to violations of normality. Therefore, minor deviations from normality are unlikely to invalidate the results

(Heeren & D'Agostino, 1987). To assess normality, the Shapiro-Wilk test was employed. This test evaluates whether the data in each group follow a normal distribution. A p value greater than .05 indicates that the data are consistent with a normal distribution, while a p value less than .05 suggests significant deviations from normality. The Shapiro-Wilk test results indicated that the assumption of normality was met for all outcome variables, as the tests were not significant ($p > .05$). Specifically, the results showed that a normal distribution was exhibited for declarative knowledge of epidemics ($W = .94$, $p = .26$), declarative knowledge of complex systems in epidemiology ($W = .91$, $p = .13$), and explanatory knowledge of complex systems in epidemiology ($W = .84$, $p = .08$). These findings confirm that the data for each dependent variable satisfy the normality requirement for conducting subsequent analyses.

The final assumption test was the assumption of homogeneity of variance. Homogeneity was tested using Levene's test for equality of variances. This test is interpreted such that if the significance value is greater than $p > .05$, the variances between groups are equal, and the assumption of homogeneity is met. Conversely, if $p < .05$, the variances are unequal, indicating a violation of the assumption (Keselman et al., 1979). The results showed that the assumption of homogeneity was met for all three outcome variables, as Levene's test was not significant in each case ($p > .05$). Specifically, equal variances across conditions were found for declarative knowledge of epidemics ($F(1, 33) = 0.067$; $p = .79$), declarative knowledge of complex systems in epidemiology ($F(1, 33) = 0.056$; $p = .81$), and for explanatory knowledge of complex systems in epidemiology ($F(1, 33) = 0.078$; $p = .78$). These findings confirm that the assumption of homogeneity of variance was satisfied for all dependent variables.

Therefore, based on the results of the assumption testing, a series of independent samples t-tests was deemed appropriate to address RQ1: "Does the PF condition lead to superior learning outcomes in *declarative* knowledge of epidemics, *declarative* knowledge of complex systems concepts in epidemiology, and *explanatory* knowledge of complex systems concepts in epidemiology, as compared to the DI condition?" To account for the increased risk of Type I error associated with conducting multiple t-tests, a Bonferroni correction was applied. This adjustment set a new alpha level of 0.0167 (α = 0.05/3 = 0.0167), meaning that *p*-values below this threshold were considered significant, whereas *p*-values exceeding it were deemed non-significant (Sedgwick, 2012). This approach ensured a more stringent control over the likelihood of falsely rejecting the null hypothesis. It was also decided that effect size would also be taken into consideration as per Cohen's (1988) thresholds, where a threshold of d = 0.50 would differentiate between a small and a medium effect. In addition, a paired-samples t-test was conducted within the two groups (PF and DI) between pre- and post-test scores to determine whether the students in the PD or DI learning condition exhibited better learning outcomes when assessing the three types of instructed knowledge.

The traditional null hypothesis significance testing was complemented with Bayesian analysis (Bayesian independent samples t-test and paired-samples t-test) with the help of JASP (version 0.19.3; JASP Team, 2025). The Bayes factors (BF_{10}) in this study were interpreted as per the criteria of Jeffreys (1961) and Kass and Raftery (1995); that is, $BF_{10} > 3$ shows substantial evidence for the alternative hypothesis (H_1), while $BF_{10} < 0.33$ exhibits substantial evidence for the null hypothesis (H_0).

Scoring of quantitative data for RQ1. Marking rubrics were adapted from Jacobson et al. (2017) to score the students' answers to the open-ended questions of both the pretest and post-test (See Appendix I). These marking rubrics consisted of a four-point coding scheme (see Table 7) with scores as follows: (a) a score of 0 was assigned for incorrect answers or irrelevant ideas, (b) a score of 1 was assigned for a partially correct response, where students used some terms appropriately but without further elaboration, (c) a score of 2 was assigned when the student provided appropriate terms with some scientifically correct explanations, and (d) a score of 3 was assigned for answers that show complete expert explanation and full understanding of the concepts. Table 7 presents examples of the scoring method for one of the questions testing declarative knowledge of epidemics. Table 8 provides examples of the scoring method for one of the questions testing declarative knowledge of complex systems concepts in epidemiology, whereas Table 9 provides examples for one of the questions testing explanatory knowledge of complex systems concepts in epidemiology.

Table 7

Rubric Examples of Declarative Knowledge of Epidemics Question, "What Is the Difference Between Latency and Incubation Period?"

Score	Answer Type	Description	Examples	Comments
NA	No answer/off task	No answer	Answer left blank, smiley, or question mark	Failed to address the questions
0	• Incorrect response • Shallow answer • Irrelevant ideas	• Scientifically irrelevant or incorrect idea • Lack of seriousness in ideas presented • Saying the same thing as in question	"They are similar stages but one comes before the first." "How a disease is tracked through computer networks."	Irrelevant or incorrect answer, showing no grasp of the concepts
1	• Partially correct or incomplete • No use of scientifically correct terminology	• Uses potentially related terms but imprecisely • Uses scientific terms correctly, but without sufficient context	Latency: "organism is in the body but not replicating." Incubation Period: "organism is replicating before symptoms appear."	Overly simplistic but partially correct, lacking key distinctions
2	• Correct answer with some/no scientifically correct terminology • Correct answer with incomplete or unclear information	• Uses two or more terms correctly • Uses some terms correctly but with incomplete or unclear connection to relevant ideas	Latency: "individual has no symptoms" Incubation period: "replicating period before symptoms" Latency: "the infectious organism does not replicate." Incubation period: "period where the infectious organism replicates."	Replication is correctly separated but critical context is omitted.
3	• Correct answer with scientifically correct terminology • Correct answer with complete and clear information	• Uses scientifically correct terms in precise way • Reasonably complete expert definition of terms showing understanding of context and connections to relevant ideas	Latency: "the time when a pathogen persists without replicating." Incubation period: "the period from the start of the infection to the time in which symptoms of the disease start appearing."	Accurate definitions

Table 8

Rubric Examples of Declarative Knowledge of Complex Systems in Epidemiology Question, "What are examples of dynamic equilibrium in epidemics?"

Score	Answer Type	Description	Examples	Comments
NA	No answer/off task	No answer	Answer left blank, smiley, or question mark	
0	• Incorrect response • Shallow answer • Irrelevant ideas	• Scientifically irrelevant or incorrect idea • Lack of seriousness in ideas presented • Saying the same thing as in question	• Vaccinations and requirements needed to avoid or live with the infection. • Epidemics aims to create a disease-free world	Answers have irrelevant content, or they are off task.
1	• Partially correct or incomplete • No use of scientifically correct terminology	• Uses potentially related terms but imprecisely • Uses scientific terms correctly, but without sufficient context	• New infections and recovery are the same	There is an indirect indication of time. It also has a limited understanding of the main concepts.
2	• Correct answer with some/no scientifically correct terminology • Correct answer with incomplete or unclear information	• Uses two or more terms correctly • Uses some terms correctly but with incomplete or unclear connection to relevant ideas	• It refers to a state when rate of new infections and rates of recovery are equal – so a stable number of people in the population have the disease.	Lacks elaboration or an example to demonstrate an understanding
3	• Correct answer with scientifically correct terminology • Correct answer with complete and clear information	• Uses scientifically correct terms in precise way • Reasonably complete expert definition of terms showing understanding of context and connections to relevant ideas	• Dynamic equilibrium in epidemics refers to a state where the rates of new infections and recoveries (or removals) are equal, resulting in a stable number of active cases over time. For example, in endemic diseases like the common cold, the total number of infected individuals relatively constant.	Accurately describe the concept and provide examples

Table 9

Rubric Examples of Explanatory Knowledge of Complex Systems in Epidemiology Question, "What are examples of dynamic equilibrium in epidemics? Please explain."

Score	Answer Type	Description	Examples	Comments
NA	No answer/off task	No answer	Answer left blank, smiley, or question mark	
0	• Incorrect response • Shallow answer • Irrelevant ideas	• Scientifically irrelevant or incorrect idea • Lack of seriousness in ideas presented • Saying the same thing as in question	Diseases spread because people get sick, and they eventually stop spreading.	No reasoning or explanation of dynamic equilibrium concept
1	• Partially correct or incomplete • No use of scientifically correct terminology	• Uses relevant terms without correct explanation • Uses terms correctly, but explained incorrectly	Diseases spread more when people are indoors, but vaccinations can stop it.	Limited reasoning and no connection to the Dynamic Equilibrium/ balance concept
2	• Correct answer with some/no scientifically correct terminology • Correct answer with incomplete explanation	• Uses two or more terms correctly • Uses some terms correctly but explanation is weak	During flu season, more people are indoors, which increases the spread of the virus. Vaccinations help reduce infections and balance the number of cases.	Correctly explain the "balance" concept but lack elaboration
3	• Correct answer with scientifically correct terminology • Correct answer with complete explanations and elaboration	• Uses scientifically correct terms • Reasonably complete expert explanation shows understanding of context, causal relations, and connections to relevant ideas	There is a balance between factors that promote the spread of a disease and those that work against it. Like during the flu season, more people are inside because of the weather which increases the rate of transmission but measures like vaccinations, hand washing help decrease the spread of the virus.	Using an example to demonstrate the understanding of the key 'balance'

The pretest, post-test, and Challenge Problem 3 Application Task data (see below) were scored by two raters. One was a senior lecturer at the University of Western Australia Medical School with a public health and epidemiology background and prior scoring experience. The other was a medical practitioner specialising in epidemiology but with no prior scoring experience. To calibrate their scoring, the two raters first scored 10% of the student responses independently and then met to discuss the discrepancies and reach agreement. This procedure was repeated with the next 10% of student responses. Finally, the remaining responses were scored independently.

To calculate inter-rater reliability (i.e., a statistical measurement of the extent of consistency or agreement between two or more observers or raters evaluating the same phenomenon; McHugh, 2012), Cohen's Kappa was used, it being the most advantageous method when having categorical data and two raters. The Kappa value ranges from 0 to 1, with values closer to 1 indicating perfect alignment between the raters (Viera & Garrett, 2005). For the scoring of the experimental group's pretest responses there was a Kappa value of 1, indicating perfect agreement between the two raters. For the experimental group's post-test scoring, there was a Kappa value of .891, indicating strong agreement between the raters. Likewise, scoring of the control group's pretest responses achieved a Kappa value of .855, indicating strong agreement between the raters, and scoring of the control group's post-test responses achieved a Kappa value of 1, indicating perfect agreement between the raters. Whenever there was a discrepancy, the final score for a response was agreed upon by the raters in conjunction with the researcher.

There was a very strong correlation between declarative knowledge of complex systems in epidemiology and explanatory knowledge of complex systems in epidemiology (r = .96). This high correlation may raise questions about whether the scoring of these constructs accurately captures the differences between declarative and explanatory knowledge.

However, previous research has identified these types of knowledge as integrated learning approaches (Jacobson et al., 2017), which may help to justify the strong relationship observed.

Importantly, there are key differences in the coding rubrics used for these constructs. The rubric for declarative knowledge emphasises recalling and stating examples or factual information, while the rubric for explanatory knowledge focuses on reasoning and articulating concepts, including causal links. Specifically, declarative knowledge assessments ask for examples, whereas explanatory assessments require more in-depth explanations. Furthermore, the evaluation criteria for the rubrics differ; declarative knowledge is assessed based on accuracy, completeness, and use of scientific terminology, whereas explanatory knowledge is evaluated based on clarity of explanation, understanding, and depth of reasoning. These distinctions suggest that the strong correlation between the two is not simply a result of overlapping coding constructs but rather reflects a genuine relationship between these integrated aspects of knowledge in epidemiology.

RQ2: Hypotheses and Method of Analysis

Hypotheses for RQ2. The second research question calls for evaluating both the PF and DI groups' ability to transfer learned knowledge to solve new problems within the same domain (near within domain transfer) and in a new domain not previously studied (far across domain transfer). Near within domain transfer was tested on the pretest and post-test by transfer problems 1 and 2, whereas far across domain transfer was tested only on the post-test by transfer problem 3. Two hypotheses were developed in relation to RQ2:

H3 (directional): The PF-treatment group will exhibit greater gains than the DI-control group on *near within* domain transfer problems.

H4 (directional): The PF-treatment group will exhibit greater gains than the DI-control group on *far across* domain transfer problems.

Choice of data analysis for RQ2. To test H3, it was decided that paired samples t-tests would be used separately for the PF and DI conditions to determine whether either group made significant gains between the pretest and post-test on the near transfer questions. In addition, an independent samples *t*-test would be performed between the two groups to determine whether their post-test scores on the near within domain transfer problems were significantly different. To test H4, independent samples *t*-tests would be used for a between-group comparison of the two groups' post-test scores on the far across domain transfer problems.

Like the RQ1 analysis, here the traditional null hypothesis significance testing was complemented with Bayesian analysis (Bayesian independent samples t-test and paired samples t-test) with the help of JASP (version 0.19.3; JASP Team, 2025). The Bayes factors (BF_{10}) were interpreted as per the criteria of Jeffreys (1961) and Kass and Raftery (1995), i.e., $BF_{10} > 3$ shows substantial evidence for H_1, and $BF_{10} < 0.33$ exhibits substantial evidence for H_0.

Scoring of quantitative data for RQ2. Table 10 shows representative responses that received scores from 0-3 for the "Common Cold" near within domain transfer problem and the "Road Injury and Mortality" far across domain transfer problem.

Table 10

Examples of Participants' Post-test Transfer Problem Solutions and Rubric Scores

Score	Near transfer example ("Common Cold" problem)	Far transfer example ("Road Injury & Mortality" problem)
0	Epidemics: "the spread of disease" Outbreak: "high increase in cases"	No participants received a 0.
1	"Outbreak when a disease happens within a community but when it spread more than expected, it's called epidemic."	"Probably one rule cannot be generalized. Also, one might take it in mind or not. putting different regulatory methods increase the chances of not falling in the error and if the person skipped one rule he will stick to the others. that's why governments try different ways and methods to enforce people to go by regulatory measures"
2	"Epidemic is the occurrence of the disease in a certain region with very high excess cases, but the outbreak refers to the appearance of a disease in a small area. According to the above cases common cold cases indicate an epidemic, while cases of tuberculosis are considered as an outbreak."	"Having different preventative measures regarding the usage of mobile phones while driving make the people more aware of the danger that the phones cause and leave the less distracted and more focused to the traffic signs and reduce the accidents and the mortality rate caused by road traffic accidents."
3	"Although the concepts can be relatively used interchangeably there is still tiny difference between the two concepts when looking to a bigger realm. Both of them share the quality of an increased number of cases of a certain disease in geographical area within the expected range. However, epidemics tend to cover large group of people with a larger geographical area than outbreak. In the other hand, outbreaks tend to be more sudden and for shorter period of time. Also, outbreaks usually encompass smaller geographical area comparing to the	"Every driver is different. Therefore, diversifying the means and methods of prevention is required to be directed to all types of drivers, thus containing the problem from all aspects and reducing the chances of accidents and distractions. Different prevention methods may target different causes, and each cause has many risks and consequences, so diversifying prevention methods has tremendous results in terms of reducing the number of injuries and deaths. This will also instil a sense of commitment among drivers and reduce the use of the phone for fear of the consequences that may result from using the phone. The methods of prevention may be varied, including financial violations that the user of the phone while driving must pay. It is also possible that one of the methods of prevention is to impound the car for a period of time with the various authorities. All of this will ultimately lead to reducing the use of the phone while driving for fear of the consequences that may result from using the phone while driving. In addition, spreading awareness, advice and guidance,

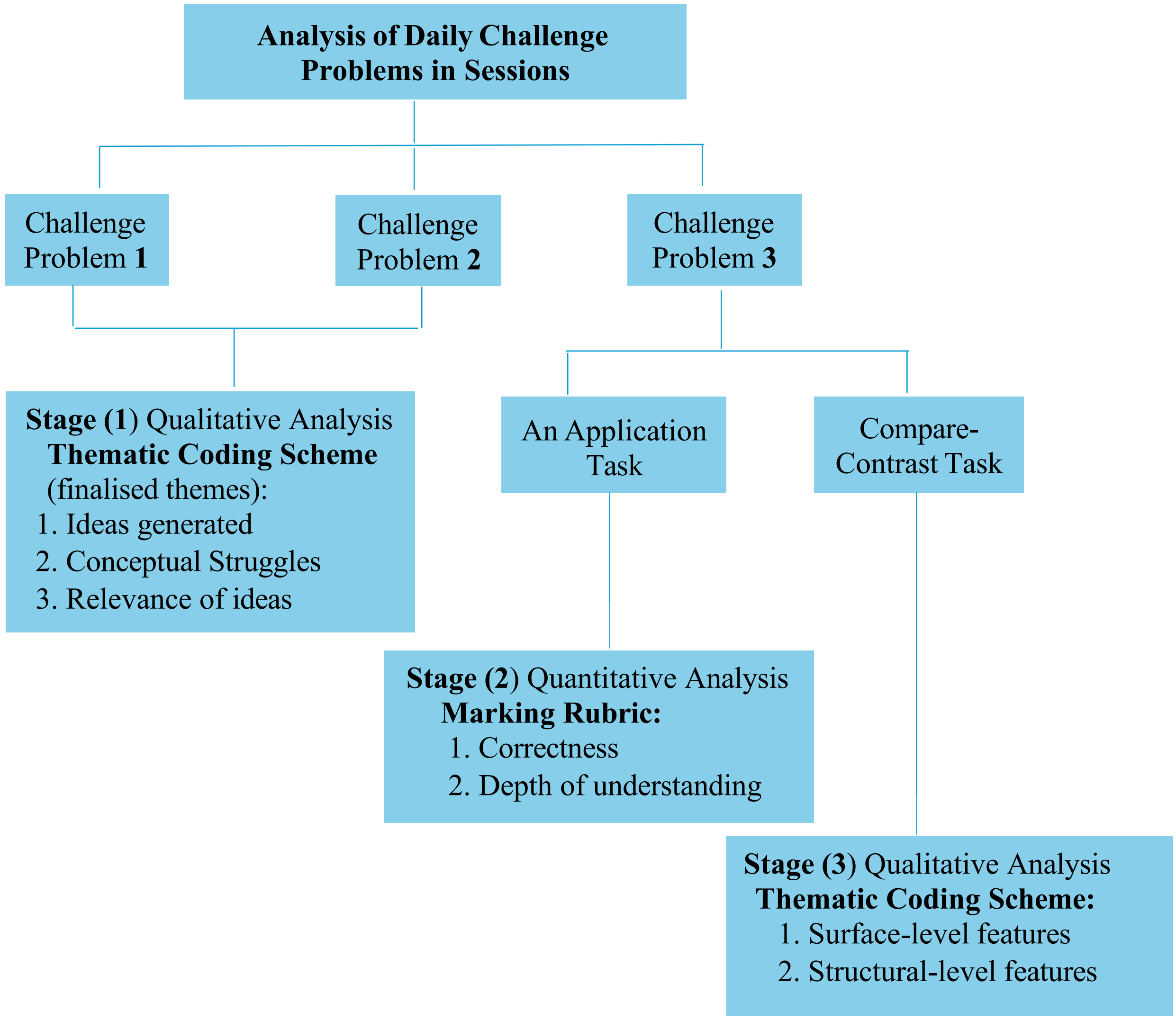
Score	Near transfer example (“Common Cold” problem)	Far transfer example (“Road Injury & Mortality” problem)
	epidemics. there is more specificity in the group of people who suffer the outbreak. for example, an outbreak could affect black-Americans more than white American because of certain type of food which is more liked by black Americans. To conclude, outbreaks can escalate if they spread to larger geographical areas and involve more diverse populations.”	in addition to informing drivers of the risks resulting from using the phone while driving, which vary between financial violations to injuries and accidents that will lead to loss of life.”

RQ3: Method of Analysis

The third research question sought to evaluate how the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts affected the learning process across multiple sessions in PF vs. DI conditions. The session- by-session challenge problems were thus analysed to further investigate the impact of the ABM-based PF vs. DI approach on the learning outcomes. The students' solutions to these problems (which were presented in some cases before and in some cases after the instructional intervention, depending on the PF or DI condition) were compared both within and between the two treatment conditions. The different nature of the challenge problems warranted different forms of analysis and different criteria, an overview of which is provided in Figure 13.

Figure 13

Overview of Analyses of Data from Challenge Problems for RQ3



Overview of analysis of Challenge Problems 1 and 2 data. In the first phase of the PF design, that is, the generation and exploration phase, students explored an ABM each session to solve an epidemics problem and then another ABM to solve a complex systems problem. They were not expected to produce correct solutions when completing these tasks; rather, the intent of this phase was to promote cognitive engagement through productive struggle, activate their schema, and generate ideas, which might include incomplete or flawed solutions.

The responses to Challenge Problems 1 and 2 were analysed using thematic analysis, a widely used qualitative research method involving a systematic process of iteratively reviewing the data to identify, evaluate, and report repeated patterns (Braun & Clarke, 2006). Thematic analysis is distinguished from other approaches by its theoretical flexibility, procedures of coding and theme development, and both deductive and inductive orientations to analysis (Braun & Clarke, 2022). In the present analysis, four broad themes were decided before data analysis based on prior research (i.e., a deductive approach) but were later narrowed to just three based on the data itself (i.e., inductive). Moreover, within each broad theme, new sub-themes emerged from the data during coding. Overall, then, the process involved a hybrid deductive/inductive approach to thematic analysis.

Coding can also be automated or performed manually. The latter requires careful reading and interpretation of the text, thus ensuring that the intent and context behind the words are duly considered. In this study, all coding and analysis of the qualitative data were performed manually, initially by the researcher and then with follow-up assistance from two external coders—a data consultant and a qualitative researcher. These external coders helped resolve ambiguous or otherwise problematic coding decisions and also co-coded a substantial portion of the data as a means of verifying the reliability of the researcher's coding (see below for details).

To maintain analytic clarity and address RQ3's focus on learning over time, data were carefully segmented for analysis:

- By Group (PF / DI)
- By Session (1, 2, 3)
- By Task (Challenge Problem 1, Challenge Problem 2)

This segmentation allowed changes over sessions to be observed clearly (e.g., whether PF students showed reduced struggle in later sessions).

Six steps of thematic analysis for Challenge Problems 1 and 2. Braun and Clarke's (2006) method of thematic analysis, applied here to ensure the rigor and transparency of the present study’s analysis, includes six steps: familiarisation with data, coding, generating themes, reviewing themes, naming and defining themes, and producing a report of the findings. In broad terms the data was iteratively reviewed, during which time codes were added (i.e., labels were assigned to segments of data), removed, refined, or merged (where needed) and organised into themes and sub-themes.

During initial thematic development, PF and DI group responses were reviewed separately to allow distinct patterns, struggles, or reasoning styles to emerge without conflating the groups to ensure the themes and sub-themes developed were not biased toward one instructional design. However, once the themes were established, the separate codebooks were merged, and the same final codebook was applied to both groups’ data. This allowed for meaningful comparison of learning outcomes across the PF and DI conditions using a consistent analytical framework.

Following is a more detailed description of how each of Braun and Clarke’s (2006) six steps of thematic analysis was carried out for the Challenge Problems 1 and 2 data.

Step 1: Familiarisation with the Data

All responses were read repeatedly by the researcher to gain familiarity with the diversity of student reasoning. Students’ reasoning varied considerably in terms of depth, from a simple “high contact rate” to a multi-variable description linking SEIRS model parameters, hygiene, population density, and social behaviour. One example of an early impression the researcher formed of differences among student responses is that some students carefully listed SEIRS variables while others mentioned more general social factors, making coding for relevance crucial.

Step 2: Generating Initial Codes

In this step, the researcher conducted manual coding by reading each line of the data, identifying important terms and concepts in the students' comments, and giving them codes or labels that provided context. These codes or labels were typically sentences or meaningful clauses reflecting a single idea, such as describing infection probability versus discussing hygiene. Other examples of raw codes include:

- “short immunity period”
- “mutation and new strains”
- “poor sanitation”
- “high population density”
- “misunderstanding latency vs. incubation”
- “technical terminology”

The qualitative analysis accounted for researcher positionality by examining how the coders’ previous knowledge and educational background in learning sciences influenced student response interpretation. To this end, throughout the coding process, a reflexive journal was kept in order to document analytic decisions along with unexpected interpretations and developing insights to reduce potential bias. Decisions made about confusing codes and quotes were also recorded in this journal. The use of the journal helped maintain uniformity and transparency across the several sessions and coders.

Step 3: Searching for Themes

In this step, the aforementioned initial codes were grouped by the researcher into broader themes, informed by both deductive and inductive reasoning. Deductively, themes were at times drawn from the research questions and theoretical framework whenever these matched the data; inductively, new ideas for themes were drawn from the data itself. This dual approach allowed for both a theory-driven focus and openness to novel insights from participants' responses.

Initially, four main themes were deductively derived from the literature, inspired in particular by Kapur (2008):

- **Breadth of Ideas Generated**: Capturing the number and variety of ideas (e.g., SEIRS parameters, social/environmental factors).
- **Relevance to Problem**: Assessing how appropriate or contextually meaningful the student’s ideas were to the given model.
- **Conceptual Struggles**: Identifying misunderstandings or misapplications of terms or concepts.
- **Technical Language Use**: Detecting precise terminology linked to model variables or epidemiological constructs.

This preselected framework aligned with the theoretical underpinnings of the PF and DI approaches: PF was expected to produce more varied ideas and higher struggle rates whereas DI was expected to yield more relevant and technically aligned responses. The alignment with these theories ensured that the resulting thematic analysis was both conceptually coherent and theoretically transferable to new datasets, countering typical critiques about subjectivity and replicability.

For an example of inductive refinement illustrating the iterative development of theme hierarchies, the theme "Mutation" was used frequently in the initial coding and was at first treated as an independent theme. Later in the coding, however, as patterns in the data emerged, it became clear that when students used the term "Mutation" they were generally referring to any one of a number of distinct mechanisms. The “Mutation” subtheme was, therefore, relabeled as "Breadth of Ideas" to convey how the term was used in the context of model parameters.

Step 4: Reviewing Themes

Once the initial grouping of codes was complete, the original four deductively-derived themes (Breadth of Ideas Generated, Relevance to Problem, Conceptual

Struggles, and Technical Language Use) were critically reviewed by the researcher together with the two external coders through a process of iterative discussions and comparison across sessions. This phase focused on ensuring that each theme:

- Represented a distinct analytical category
- Was applied consistently across PF and DI groups
- Could accommodate new sub-patterns emerging from the data

As part of this process, all coded data segments were revisited, and thematic definitions were refined. Sub-themes were also identified within each major theme to capture nuances specific to particular sessions or models. This iterative review laid the foundation for a final theme structure that could be consistently applied across challenge problem sessions and model types.

Examples of thematic refinements made at this stage include the following:

- Merging “Technical Language Use” into a broader category where its role was more explanatory (e.g., subsumed under "Relevance to Problem" or used to assess clarity).
- Replacing “Breadth of Ideas” with a more neutral theme: “Number and Variety of Ideas Generated,” which allowed the analysis to more clearly capture variation in idea richness without implying correctness.
- Clarifying the “Conceptual Struggles” theme by separating conceptual misunderstandings**,** model interpretation errors**,** and reasoning flaws into dedicated sub-themes (e.g., misunderstanding incubation vs. immunity, misreading simulation visuals).
- Refining “Relevance to Problem” to focus strictly on alignment between student responses and core model mechanics or problem demands.

To ensure analytic reliability, ambiguous responses were discussed among the multiple coders. This helped verify that disagreements often stemmed from definitional overlaps, particularly between “Relevance to Problem” and “Conceptual Struggles,” and ensured that themes were given clear operational boundaries.

For example, one student’s response, “Fire spreads because everything is close together and dry,” was originally categorised under Breadth and Technical Language. Upon review, it was more appropriately recoded as **“**Number and Variety of Ideas Generated” under sub-theme “Environmental & External Influences” within the main theme “Number and Variety of Ideas Generated.” It was also double coded (i.e., simultaneously coded) under the main theme “Relevance to Problem” given its alignment with the model's fire spread mechanism. Such double coding was used where appropriate to reflect multi-dimensional aspects of students’ reasoning.

Step 5: Defining and Naming Themes

Following the review process, the final thematic structure was established to enhance clarity, comparability, and alignment with the study’s research questions. The codebook was revised to include three final major themes, each with clearly defined sub-themes based on the content and structure of student responses (see Table 11). These themes were consistently applied across all sessions and both instructional conditions.

Table 11

Final Themes of Challenge Problems 1 and 2 Data

Final Theme	Definition	Representative Sub-Themes
Number and Variety of Ideas Generated	Captures the diversity and richness of ideas students proposed in response to the model-based tasks, including domain-specific and cross-domain mechanisms.	Biological & Epidemiological Factors, Human Behaviour & Social Factors, Mathematical & Simulation-Based Insights, Fire Spread Analogies, Adoption & Market Penetration, Predator-Prey Interactions
Conceptual Struggles	Encompasses misunderstandings, misinterpretations, or flawed reasoning evident in students' written responses.	Immunity vs. Reinfection, Confusion about Latency vs. Incubation, Misinterpretation of Visual Outputs, Overgeneralisation of Social Trends
Relevance of Ideas to the Problem	Assesses the degree to which students' ideas were appropriate and aligned with the underlying mechanics of the models used in the challenge problems.	Understanding of Core Model Mechanics, Appropriateness of Analogies & Real-World Connections

Each theme was associated with session-specific sub-themes reflecting the ABMs presented and the nature of the student responses. Examples include:

- **For Theme 1 (Ideas Generated)**: sub-themes such as *Environmental Factors* in the Fire Model and *Tipping Points & Longevity* in the Marketing Model.
- **For Theme 2 (Conceptual Struggles)**: sub-themes like *Model Interpretation Struggles* in Forest Fire and *Population Balance Misconceptions* in the Wolf-Sheep Model.

- **For Theme 3 (Relevance of Ideas)**: sub-themes tracking whether students' analogies and explanations directly related to the functional logic of the ABMs.

A more complete summary of subthemes derived for each of the three finalised main themes, along with representative supporting quotations, is provided in Tables 12-14.

Step 6: Producing the Final Analysis

The final phase of thematic analysis involved the researcher synthesising these themes into a coherent narrative in a way that tied them back to the research question.

Table 12

Summary of Theme 1: Number and Variety of Ideas Generated

Session	Challenge Problems	Sub-Themes	PF (Experimental) Group Example Quote	DI (Control) Group Example Quote
Session 1: Epidemic & Forest Fire Models	Challenge Problem 1 (SEIRS Model – Epidemic Spread)	1) Biological & Epidemiological Factors, 2) Human Behaviour & Social Factors, 3) Mathematical & Simulation-Based Insights	*"The disease spreads when immunity is not permanent. Infected who recover can get sick again."*	*"Some people don't present any symptoms of the disease, so they transmit it to others."*
	Challenge Problem 2 (Forest Fire Model – Emergent Behaviour)	1) Fire Spread Analogies, 2) Mathematical & Simulation-Based Insights, 3) Environmental & External Influences	*"The fire here follows the same concept as disease spread. The disease does not move from one person to another, but spreads from the host, giving many new sources/hosts."*	*"The spread of fire depends on how close the trees are, just like the spread of disease depends on population density."*
Session 2: Malaria & Marketing Models	Challenge Problem 1 (Malaria Model – Disease Transmission)	1) Disease Transmission Mechanisms, 2) Environmental & Population Factors, 3) Prevention & Control Strategies	*"Malaria spreads through the bite of infected female mosquitoes."*	*"If a mosquito bites infected person, it can transmit the parasite to a healthy person."*
	Challenge Problem 2 (Marketing Model – Consumer Behaviour)	1) Adoption & Market Penetration, 2) Psychological & Social Influences, 3) Tipping Points & Product Longevity	*"People are more likely to buy a product if they see others using it."*	*"Advertisement and social influence help products spread faster."*
Session 3: COVID-19 & Wolf-Sheep Models	Challenge Problem 1 (COVID-19 Model – Disease Spread)	1) Direct & Indirect Transmission Pathways, 2) Public Health Policies & Behavioural Changes, 3) Long-Term Immunity & Reinfection Cycles	*"COVID-19 spreads through respiratory or sneezing and contact with infected places."*	*"A person can be infected without symptoms and without knowing, they spread the disease."*
	Challenge Problem 2 (Wolf-Sheep Model – Population Dynamics)	1) Predator-Prey Interactions, 2) Ecological Equilibrium & Resource Availability, 3) Impact of External Factors	*"Wolves eat sheep, and when there are too many wolves, they run out of food."*	*"If wolves eat too many sheep, they will finally starve."*

*See a detailed table with descriptions of the sub-themes in Appendix J

Table 13

Summary of Theme 2: Conceptual Struggles

Session	Sub-Themes	Example PF (Experimental) Group Quote	Example DI (Control) Group Quote
Session 1: Epidemic & Forest Fire Models	1) Conceptual Struggles (Immunity, reinfection misconceptions), 2) Perceptual & Model Interpretation Struggles (Visual illusions in fire model), 3) Reasoning & Explanation Struggles (Overgeneralisations, lack of causal connections)	*"The disease spreads in waves because people get infected, recover, then lose immunity, and get infected again forever."*	*"Immunity does not stop the disease, as even immune people can get sick again immediately."*
Session 2: Malaria & Marketing Models	1) Conceptual Struggles (Misunderstanding of vector-borne transmission), 2) Perceptual & Model Interpretation Struggles (Misinterpretation of tipping points), 3) Reasoning & Explanation Struggles (Failure to connect Socioeconomic & behavioural patterns)	*"Malaria spreads from person to person just like COVID-19."*	*"People can catch malaria directly if they are near an infected person."*
Session 3: COVID-19 & Wolf-Sheep Models	1) Conceptual Struggles (Difficulty in understanding immunity & virus evolution), 2) Perceptual & Model Interpretation Struggles (Population balance misconceptions in Wolf-Sheep model), 3) Reasoning & Explanation Struggles (Failure to explain tipping points in disease & predator-prey interactions)	*"COVID-19 spreads forever because people never develop full immunity."*	*"The virus mutates immediately, so vaccines are useless."*

*See a detailed table with descriptions of the sub-themes in Appendix J

Table 14

Summary of Theme 3: Relevance of Ideas to the Problem

Session	Sub-Themes	Example PF (Experimental) Group Quote	Example DI (Control) Group Quote
Session 1: SEIRS & Fire Models	1) Understanding of Core Model Mechanics, 2) Appropriateness of Analogies & Real-World Connections	*"Fire doesn't move, but it spreads by changing tree states, like infection spreads by changing immune status."*	*"The disease spread is modelled by probabilities, just like the fire simulation."*
Session 2: Malaria & Marketing Models	1) Understanding of Core Model Mechanics, 2) Appropriateness of Analogies & Real-World Connections	*"Mosquitoes spread malaria, so preventing mosquito bites reduces infections."*	*"The tipping point in malaria spread is when too many mosquitoes are infected."*
Session 3: COVID-19 & Wolf-Sheep Models	1) Understanding of Core Model Mechanics, 2) Appropriateness of Analogies & Real-World Connections	*"If people lose immunity, COVID-19 can come back, just like wolves return when there are more sheep."*	*"The predator-prey balance works like a controlled epidemic cycle."*

*See a detailed table with descriptions of the sub-themes and key insights in Appendix J

Inter-Rater Reliability of Coding of Data from Challenge Problems 1 and 2. Inter-rater reliability for coding of this qualitative data was addressed through a consensus-building process. To corroborate the coding of the researcher, a stratified random sampling of 20% of the responses from each qualitative dataset (i.e., the Challenge Problems 1 and 2 responses considered above as well as the Compare-Contrast Task responses of Challenge Problem 3 and the Self-Reports and Focus Group transcripts treated below) were independently coded by the two external coders. This sampling approach enabled consistent comparison across subgroups while maintaining methodological rigor. The selection of samples from the Challenge Problems 1 and 2 dataset was stratified to ensure representation across sessions, instructional conditions, and task types. Rather than calculating a formal Kappa coefficient, then, the coding team adopted a negotiated agreement approach suitable for exploratory qualitative research. This iterative and collaborative process ensured the credibility and transparency of the analysis, while promoting consistent use of the thematic framework throughout the dataset.

While the main themes were theory-driven (deductive), all coders used an open and iterative approach for identifying sub-themes. These sub-themes were inductively derived from recurring patterns in the data during coding discussions. For example, “Conceptual Struggles” was a pre-established theme, but its specific sub-themes such as confusion about immunity, incubation, or tipping points were identified based on students’ varied interpretations. Likewise, “Number and Variety of Ideas Generated” was a deductively derived category, but its sub-themes (e.g., epidemiological, behavioural, environmental factors) emerged inductively across datasets. Differences in interpretation were discussed in review meetings. These discussions refined the operational definitions of both themes and sub-themes, clarified boundaries between overlapping categories (e.g., surface vs. structural features), and contributed to improved consistency in coding. A summary of this consensus-building process is provided in Table 15.

Table 15

Inter-Rater Reliability of Challenge Problems 1 and 2 Data Coding

Dataset Segment	Level of Agreement	Key Discrepancies	Resolution Approach	Outcome
Challenge Problems 1 & 2 Responses (20% subset) – Researcher vs. Coder A	Moderate-to-high agreement	Disagreements on coding “Struggles” when students showed misunderstanding vs. off-topic ideas; minor differences in sub-theming of “Types of Ideas”	Joint review of misclassified quotes; clarified definitions of “Struggle” (focused on conceptual and reasoning difficulties); added boundary examples for sub-themes	Codebook definitions refined; improved clarity in applying “Struggle” vs. “Irrelevant” codes; consistent application of sub-themes for Types of Ideas
Challenge Problems 1 & 2 Responses (20% subset) – Researcher vs. Coder B	High agreement on main themes	Overlap in coding “Breadth of Ideas” and “Relevance of Ideas” in some responses	Reviewed ambiguous cases; added sub-criteria to separate idea generation (quantity) from idea relevance (quality)	Codebook updated with clearer distinctions; better guidance on coding complex student responses
Challenge Problems 1 & 2 Responses (20% subset) – Researcher vs. Both Coders	General alignment across all themes	Occasional confusion between “Human Behaviour” and “Environmental Factors” in “Types of Ideas” theme	Group discussion of overlapping cases; refined sub-theme definitions with boundary criteria	Improved consistency in theme application; codebook updated with boundary examples; enhanced reliability for analysing differences between PF and DI groups

Blindness to Condition. To ensure the credibility and consistency of the qualitative analysis, a systematic approach was implemented for the first two qualitative datasets to ensure coder blindness to condition. This includes the Challenge Problems 1 and 2 response data addressed above and the Compare-Contrast Task data from Challenge Problem 3 addressed below. (The Self-Reports and Focus Group data were intentionally not coded blind to condition; see the relevant discussion in that section.) Coders were instructed to apply codes strictly based on the content of the student responses, regardless of group assignment. In most instances, blinding to experimental condition (PF or DI) was preserved during the initial coding stages. However, in a small number of cases, students explicitly referenced their instructional context in their responses (e.g., “we figured it out first” or “the video helped”), which made full blinding impossible. To address this, coders were reminded to base their decisions only on what was stated, not inferred, and to consistently apply codes across both conditions. A coding decision journal was also kept by all coders to record coding decisions, especially those involving ambiguity, disagreement, or sub-theme evolution.

Overview of analysis for Challenge Problem 3 data. In the second phase of the Productive Failure (PF) design, termed “Consolidation and Knowledge Assembly,” students engaged with instructional content through videos focusing on complex systems and epidemics concepts. Following this, they were presented with Challenge Problem 3, which included an application task grounded in a real-world based scenario and a compare-contrast task requiring the students to identify similarities and differences between the two ABMs used earlier in the session. Both tasks were designed to reinforce the application of the acquired knowledge, promoting deeper conceptual understanding and, consequently, enhancing students’ ability to transfer their learning across domains.

***Application task*.** The design of Challenge Problems 1 and 2 was intended to engage students in exploring the ABMs and generating multiple solutions, with explicit guidance that there were no right answers. These problems were structured to encourage cognitive activation and struggle by students in ideas generation. In contrast, Challenge Problem 3 required students to provide correct answers after receiving targeted instruction. Therefore, a marking rubric was adapted from Jacobson et al. (2017) to score the students' answers to the application task element of Challenge Problem 3 presented in Sessions 1-3 (See Appendix K). This marking rubric primarily evaluated the correctness and depth of understanding exhibited in students' responses. It consisted of a four-point rating scheme with scores as follows: (a) a score of 0 was assigned for incorrect answers or irrelevant ideas, (b) a score of 1 was assigned for a partially correct response, where students used some terms appropriately but without further elaboration, (c) a score of 2 was assigned when the student provided appropriate terms with some scientifically correct explanations, and (d) a score of 3 was assigned for answers that show complete explanation and full understanding of the concepts. Table 16 summarises the ratings used and shows representative responses that received scores from 0-3 for the application task in Session 1.

Table 16

Rubric Examples for the Application Task of Problem 3 in Session 1: "Provide the Minister a summary of your assessment on how the disease is spread amongst the population depending on the disease characteristics".

Score	Answer Type	Description	Examples	Comments
NA	No answer/off task	No answer	Answer left blank or unrelated response.	
0	• Incorrect response • Shallow answer • Irrelevant ideas	• Scientifically irrelevant or incorrect idea • Lack of seriousness in ideas presented • Saying the same thing as in question	The disease spreads through a combination of microscopic changes that are unknown and coincidences.	Just saying words that aren't related to the question.
1	• Partially correct or incomplete • No use of scientifically correct terminology	• Uses relevant terms without correct explanation • Uses terms correctly, but explained incorrectly	The disease spreads due to a mix of factors, including transmission through close contact and environmental conditions, but further details are needed to fully understand its dynamics.	Starting to convey the idea, but there is not enough to determine completeness.
2	• Correct answer with some/no scientifically correct terminology • Correct answer with incomplete explanation	• Uses two or more terms correctly • Uses some terms correctly but explanation is weak	The disease can spread when people are in close contact with each other and more so if they are in a crowded space.	The idea is taking shape, but it's not fully there yet.

Score	Answer Type	Description	Examples	Comments
3	• Correct answer with scientifically correct terminology • Correct answer with complete explanations and elaboration	• Uses scientifically correct terms • Reasonably complete expert explanation (shows understanding and connections to relevant ideas)	We are looking at a disease that could be spreading from person to person. This could be via direct or indirect contact. The dangers associated with this form of spread is that if it is very virulent it could spread quickly through regular interactions. If it is airborne transmitted, which could be from droplets or the particles from droplets. This would mean sharing the same space as someone infected could result in transmission. This would mean it would be hard to identify who gets infected as it would take a brief interaction or passing to spread.	Provides a clear understanding that the phenomenon has to do with transmission types. They do not have to provide all the examples but should be confident showing at least one.

This application task was scored by the same two raters with epidemiological backgrounds who scored the pretest and post-test responses, following a similar scoring process to that used to mark those data. To test inter-rater reliability, Cohen's Kappa was used; its values range from 0 to 1. Values closer to 1 shows perfect alignment between the raters. The test revealed a perfect score of 1.00 for the scoring of both experimental and control group responses (i.e., perfect agreement between the raters).

The scored data from the application task of Challenge Problem 3 was analysed using a statistical test, unlike Challenge Problems 1 and 2, which were assessed using a qualitative approach based on thematic analysis. The first reason for conducting a quantitative analysis of the application task is that it was intended to answer the question "whether" the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts affected the learning process in each of multiple sessions, as compared to the "how" part of the question, which would require in-depth qualitative analysis. Secondly, the structure of Challenge Problems 1 and 2 entails qualitative analysis since these required students to generate ideas without regard to there being a "right" answer. In contrast, in the application task of Challenge Problem 3 the question was straightforward, and students were expected to write correct answers after receiving the instruction. Therefore, the data from this task was eligible to be subjected to statistical analysis, assessing the correctness and the depth of understanding based on the rubric.

To compare whether there were statistically significant differences in the scores of the application task of Challenge problem 3 in each session between the experimental and control groups, it was decided that a Mann-Whitney U test would be performed. This test is a non-parametric alternative to independent samples *t*-test. As shown in Table 17, the Shapiro-Wilk test was applied to the data and rejected the null hypothesis of 'normal distribution' ($p < 0.05$);

therefore, the Mann-Whitney U test was considered appropriate to run and determine any significant differences in the rating scores between the PF and DI groups.

Table 17

Tests of Normality for Challenge Problem 3 Data

Group		Shapiro-Wilk		
		Statistic	df	Sig.
Session 1	Experimental	0.865	20	0.010
	Control	0.896	15	0.082
Session 2	Experimental	0.736	20	0.000
	Control	0.842	15	0.014
Session 3	Experimental	0.760	20	0.000
	Control	0.860	15	0.024
All Sessions	Experimental	0.924	20	0.121
	Control	0.942	15	0.410

For the scores on the application task of Challenge Problem 3 across all three sessions, it was decided that an independent sample *t*-test would be run, which, similar to Mann-Whitney, examines significant differences in the scores between the two groups. The assumption of normality is upheld for both groups ($p > 0.05$), justifying the use of a parametric test, specifically an independent samples t-test.

Compare-Contrast Task. As discussed in Chapter Two, the AC has been aligned with the PF design as mutually supporting knowledge transfer by helping students abstract the underlying principles that can be applied across domains. In this study, after exposure to instruction, examples, and canonical solutions in the instructional videos, students were given a task to identify similarities and differences between the two models they had explored earlier in each session. However, it should be noted that, unlike the formal construal of AC, the compare-contrast task utilized here provided no structured prompts or explicit scaffolding that might help students directly compare the two models and extract deeper analogies. Also, as this study was conducted online asynchronously, neither was there any face-to-face discussion initiated with students to compare their attempted solutions with the canonical ones after instruction. Instead, students simply observed the examples of canonical solutions through the instructional videos and then were asked to solve the comparison tasks.

The features of the Compare-Contrast Task element of Challenge Problem 3 are summarised in Table 18.

Table 18

Compare-Contrast Task of Challenge Problem 3 in Each Session

Session	Compare-Contrast Task	Description
1	Write down what you think are the main similarities and differences between the two models you have looked at.	Students were asked to identify the similarities and differences between the SEIRS and Forest Fire Models they had explored earlier.
2	Please explain how the concept of tipping point applies to the two models you have investigated.	Students were required to apply the concept of "Tipping Points" to both the Malaria and Marketing Models, and identify the shared principles.
3	Write down what you think are the main similarities and differences between the two models you have looked at.	Students were asked to identify the similarities and differences between the COVID-19 and Wolf-Sheep Predation Models they had explored earlier.

For the analysis of the Compare-Contrast Task dataset, the researcher applied a framework drawn from the work of Gentner et al. (2003) and Jacobson et al. (2020). This framework categorises features into Surface-Level Features (SF), which highlight readily apparent characteristics, and Structural-Level Features (ST), which delve into the underlying organisation.

- Surface-Level Features refer to similarities/differences that are observable, literal, or superficial (e.g., entities used, visual outputs, terminology).
- Structural-Level Features capture causal, dynamic, emergent, and systemic relationships between model components (e.g., feedback loops, self-sustaining spread, parameter interplay).

This framework was appropriate for the present dataset because the students' task involved comparing and contrasting two simulation models during each session (e.g., SEIRS against Forest Fire, Malaria versus Marketing, and COVID-19 versus Predator-Prey). The framework helped the researcher assess not just *what* students compared but *how deeply* they understood the underlying system dynamics.

To execute the above coding scheme, a thematic analysis technique was used (Braun & Clarke, 2006). The thematic coding and analysis of the Compare-Contrast Task data under consideration here were carried out according to the same six-step manual process used to code the qualitative data from Challenge Problems 1 and 2 (based on Braun & Clarke, 2006). As noted above, the two broad themes, Surface-Level Features and Structural-Level Features, were decided based on prior research before the data analysis (i.e., a deductive approach). However, as coding progressed, new sub-themes emerged organically within each main category. For example, the Structural-Level category expanded to include sub-themes such as Emergent Properties, Transmission and Repeatability, and Thresholds and Critical Mass. Likewise, the

Surface-Level category was refined to feature Shared Visual Features, Observable Triggers, and Differences between Entities. Through ongoing revisions carried out with the input of the same two external coders who assisted in the analysis of the dataset from Challenge Problems 1 and 2, the team updated the codebook to better mirror the detail in student responses. A particular challenge was deciding how to classify comments about "ICU capacity." The team agreed that if a student described ICU capacity in terms of system-level tipping points or thresholds, it should be coded as structural. Otherwise, it was recorded as surface-level factual recall. Note that each individual student response served as the unit of analysis. When a response included multiple comparisons, it was divided into separate idea units. For instance, if a student commented on both the visualisation and transmission dynamics within one answer, each aspect was coded independently in the relevant categories.

Steps were taken to keep coders blind to student condition (PF vs. DI) during the initial thematic coding in order to avoid bias in categorising responses as structurally or superficially grounded. (Refer to the discussion of blindness to condition in regard to Challenge Problems 1 and 2 data for more details and discussion of how exceptional cases were handled.) Only after finalising the themes and sub-themes were the codes compared across conditions for analysis.

The resulting codes were organised into the themes and sub-themes shown in Table 19. Again, the same coding scheme was applied to both the experimental and control groups.

Table 19

Summary of Compare-Contrast Task Themes

Session	Theme	Sub-Themes	Example Control Group Quotes	Example Experimental Group Quotes
Session 1	Structural-Level Features Description: *Identifying deeper structural aspects*	1) Emergent Properties, 2) Factors Influencing Spread, 3) Model Complexity & Outcomes, 4) Transmission & Repeatability, 5) Immunity Differences	*"They both look at emergent properties of something that spreads."*	*"Both models are examples of complex systems. They demonstrate emergences and contain important parameters that control the microlevel of the system."*
	Surface-Level Features Description: *Identifying similarities or differences that are visible and superficial*	Shared Visual Features	*"The spread of fire in the forest is like the spread of disease among people."*	*"Both give a visual representation of the spread. In both models, we can adjust some variables."*
Session 2	Structural-Level Features Description: *Explain deeper causal, dynamic, and emergent patterns that drive tipping points in the models.*	1) Impact of External Conditions, 2) Role of Preventive Measures, 3) Self-Sustaining Spread, 4) Thresholds and Critical Mass	*"Following the policies of prevention of malaria will help to reduce the risk of malaria."*	*"In this parameter: Paid media investment fraction: A tipping point occur when the investment fraction rise to a critical point, and as a result increase customer engagement and sales."*
	Surface-Level Features Description: *Capture observable effects and measurable indicators of tipping points in the models*	1) Key Drivers of Tipping Points, 2) Observable Triggers of Tipping Points, 3) Rapid Spread & Exponential Growth	*"Tipping points can be seen if the infection period takes longer than usual and requires stronger control policies."*	*"Tipping points may occur when the adoption fraction of a product reaches a critical level."*
Session 3	Structural-Level Features Description: *Explain deeper causal, dynamic, and emergent patterns shaping system behaviour.*	1) Control Mechanisms, 2) Model Complexity & System Dynamics, 3) Resource Dependency & Population Balance, 4) Equilibrium & Stability Mechanisms	*"Both systems show connections within their populations."*	*"In both cases, the behavior of individuals or sheep and wolves impacts the dynamics of the overall system."*
	Surface-Level Features Description: *Capture observable effects, visual representations, and measurable system changes.*	1) Graphical & Visual Representation, 2) Spatial & Environmental Configuration, 3) Nature of Entities & Domain Context	*"Both show graphical representation."*	*"Both show graphical representation."*

*See a detailed table with descriptions of the sub-themes and key insights in Appendix J

As with the qualitative dataset from Challenge Problems 1 and 2, normal inter-rater reliability measures, such as Cohen's Kappa, were not calculated for this Compare-Contrast Task dataset. Instead, as before, the reliability of the coding of this dataset was evaluated by having the two external qualitative coders co-code a stratified random subset of 20% of the responses (i.e., stratified by group condition). The coders reviewed their coding independently, discussed any discrepancies openly, and refined the definitions in the codebook to enhance clarity and consistency.

Disagreements usually occurred in borderline situations, for instance, when surface-level descriptions suggested underlying structural features without explicitly stating them. These cases were resolved by jointly revisiting representative quotes and establishing clear criteria for sub-theme classification. Table 20 summarises the inter-rater reliability results for this dataset.

Table 20

Inter-Rater Reliability of Compare-Contrast Task Coding from Challenge Problem 3

Dataset Segment	Level of Agreement	Key Discrepancies	Resolution Approach	Outcome
Compare-Contrast Task (20% subset) – Researcher vs. Coder A	High agreement with nuances	Overlap in coding “Model Complexity” vs. “Control Mechanisms”; some confusion with domain vs. structure	Reviewed codebook distinctions between structural feedback and complexity of agents/rules	Definitions clarified for system dynamics vs. policy-influenced complexity; examples updated
Compare-Contrast Task (20% subset) – Researcher vs. Coder B	Substantial agreement	Difficulty in distinguishing “Resource Dependency” from “Population Balance” in wolf-sheep scenarios	Joint review of example quotes; refined criteria for coding resource thresholds vs. population feedback loops	Codebook updated with separate identifiers for ecological vs. human resource mechanisms
Compare-Contrast Task (20% subset) – Researcher vs. Both Coders	General alignment	Variation in interpreting “System Interdependence” vs. “Emergent Properties”	Conducted cross-coder discussion and merged overlapping codes into a clarified definition of emergence	Final themes adjusted to reflect micro-macro linkage and reinforced feedback mechanisms across models

RQ4: Method of Analysis

The fourth RQ focuses on student perceptions of their own learning experiences, as these may differ between the PF and DI groups. The data included self-reports by students which were generated at the end of each content session and focus group interviews. Being qualitative in nature, such data provides rich and detailed descriptions as well as reflects the interpretations and perspectives of the participants.

Overview of analysis of self-reports and focus group interviews. These transcripts were anonymised; however, for identification purposes and before proceeding with coding of the data, each student statement was categorised as experimental or control (for the student's group membership) and as male or female (for the student's sex). This means that for the RQ4 dataset (self-report and focus group responses), coding was not blind to condition. Coders were deliberately informed about whether responses came from PF or DI groups, as data were gathered and structured by instructional condition, and this segmentation was maintained during analysis. This was appropriate to the study design, given that the research questions explicitly compared student experiences across the two instructional approaches. Recognising the instructional condition enabled coders to pick up group-specific experiences, language, and thoughts that were necessary for effective comparison.

To maintain rigor and reduce bias in theme development despite this non-blind coding, the following strategies were employed:

- Predefined, theory-driven coding categories grounded in instructional theory (e.g., productive struggle in PF; clarity and efficiency in DI) guided initial coding.
- Coders focused on evidence in the text itself rather than inferring intentions or making assumptions beyond what students reported.
- Coders worked together to develop themes, check example quotes, and ensure that definitions were operationalised in consistent ways for both groups.
- Joint review sessions also considered potential over-interpretation and maintained analytic transparency.

Although awareness of condition was unavoidable and integral to answering the comparative research questions, the process emphasised consistency, clear documentation of decisions, and alignment with the study's theoretical framework. This approach ensured that thematic differences identified between PF and DI groups were grounded in participant data and could inform instructional design recommendations to support the trustworthiness and credibility of the findings.

The transcripts were analysed using thematic analysis, again based on Braun and Clarke's six-step approach (2022) described earlier. In short, the data were initially open-coded by the researcher, who looked for and labelled any aspect of the transcripts that appeared relevant to the research question. The data were then iteratively reviewed, and codes were added, removed, and/or refined, then organised into themes with the input of the two external coders in the same general manner as described with the previous qualitative datasets. As noted earlier, coding to some extent relies on whether the themes are more theory- or data-driven (Braun & Clarke, 2006). Here, the researcher approached coding primarily with the specific research question in mind (theory-driven approach). However, an inductive, data-driven approach was also considered to identify any emerging themes.

Detailed description of the six steps. A more detailed description of each step of the thematic coding of the self-reports and focus-group interview transcripts is as follows:

Step 1: Familiarisation with the data

In the first phase of the analysis, the researcher went through the data sets from the PF and DI groups multiple times. This thorough reading aided comprehension of the students' reasoning and experiences in regard to each teaching approach. By focusing on each group separately, the researcher was able to identify patterns unique to each group's experiences. For example, one PF participant noted, "I liked being able to test different scenarios on my own, which helped me think more critically." After multiple passes through the material and by cross-checking students' comments, it became clear that this and similar comments in the PF group conveyed the value students found in being able to experiment with and manipulate parameters in the ABMs, an experience unique to the PF group.

Step 2: Generating Initial Codes

Once the researcher familiarised herself with the data, she began coding each set of responses from both DI and PF groups separately. During this phase, she extracted key phrases and ideas from students' feedback and assigned descriptive labels or codes to them. For example, based on the PF data from the student self-reports and focus-group discussions, the researcher established the following codes, among others:

- Exploratory learning and experimentation (based on such comments as, “Trying out different settings gave me insights I wouldn’t get from just reading.”)
- Technical difficulties (based on comments such as, “Sometimes the simulation lagged, which made it hard to follow the steps.”)
- Effectiveness of video demonstrations (based on comments such as, “The videos really clarified how to use the models.”)

In contrast, codes such as the following were initially derived for the DI-group responses:

- Clarity and comprehensiveness of content (based on comments such as, “The instructions were clear and helped me understand the topic.”)
- Limited control over simulations (based on comments such as, “I didn’t get to explore much on my own, which felt restrictive.”)

The researcher maintained a coding journal for each dataset. This helped her to keep track of the progress and refine the codes as new patterns emerged.

Step 3: Searching for Themes

After coding the data, the researcher grouped related codes into broader themes that captured deeper meanings, being careful to keep the codes for the PF and DI data separate. One broad theme was titled "Engagement and Positive Learning Experiences." This theme included codes related to students' motivation, enjoyment, and active participation in learning. Importantly, subthemes such as "exploratory learning," "developing critical thinking skills," "the use of visual aids," and "a user-friendly website" were included only when students linked them to positive or motivating experiences. For instance, although "use of visual aids" might seem descriptive, students commonly described these aids as helping them better understand complex models, thereby enhancing engagement and fostering a sense of discovery consistent with the focus of this theme. For example, one student stated:

"The graphs really helped me see how the model changed. It made me want to try different scenarios."

Thus, all subthemes under this category were justified by their contribution to students' emotional, cognitive, or motivational engagement with the learning environment.

Another broad theme, "Challenges or Barriers to Effective Learning," encompassed sub-themes like "complexity," "technical difficulties," and "insufficient instructor guidance." These themes emerged where students expressed frustration, confusion, or a sense of limitation in interacting with the learning tools or understanding content.

Other distinct themes included "Clarity and Comprehensiveness of Content" (mostly from DI students), "Comparison of Learning Methods," and "Suggestions for Improvement."

Step 4: Reviewing Themes

The researcher together with the two external qualitative coders carefully reviewed the themes within each dataset, ensuring that they were coherent, distinct, and closely aligned with the research question. This review involved multiple rounds of comparison between PF and DI responses to confirm whether themes overlapped in intent or meaning.

For example, two closely related subthemes, i.e., “exploratory learning” from the PF group and “trial-and-error with parameters” from the DI group were initially coded separately but shared a common core: both described active engagement with model simulations. After discussion with the two external coders, these were merged into the unified subtheme “Exploratory Learning and Experimentation” which incorporated student comments such as “It was fun to see how changing the birth rate changed everything—I didn’t expect that” (from a student in the PF group) and “I tried adjusting the settings even though we weren’t told to it; helped me understand how the model worked” (DI group).

Another refinement occurred in the theme “Challenges or Barriers to Effective Learning.” Student references to software crashes and lagging simulations were combined into a subtheme called “Technical Difficulties,” which clarified that technological disruptions, not content misunderstandings, were the source of the barrier. Example statements include “The program crashed while I was answering the last quiz" and "Sometimes the simulation lagged, which made it hard to follow the steps."

This theme review also identified the need for minimal overlap between "Clarity and Comprehensiveness of Content" and engagement-based themes by separating statements about understanding material from those about emotional or motivational responses. For instance, a DI student stated, "The instructions were clear and helped me understand the topic." This response was coded under Clarity and Comprehensiveness, not under Engagement.

This step ensured that all final themes reflected clear, non-redundant dimensions of student experience while staying aligned with the underlying instructional design comparisons.

Step 5: Defining and Naming Themes

Finally, clear definitions and names were assigned by the researcher, in consultation with the external coders, to each theme so as to reflect their core significance. The themes that were finalised for the dataset based on student self-reports and focus-group transcripts are as follows:

- Engagement and Positive Learning Experiences: This captures students' appreciation for hands-on experimentation, the impact of videos and visual aids, flexibility in manipulating models, and the growth of critical thinking skills.
- Challenges or Barriers to Effective Learning: This includes technical issues, perceived complexity, insufficient guidance, and limited control over simulations.
- Clarity and Comprehensiveness of Content (mainly in DI): This

highlights students' positive views on the structured, clear instructional materials that aided in their understanding.

- Comparison of Learning Methods: This theme reflects students' thoughts on the benefits and drawbacks of computational simulations compared to traditional instruction.
- Suggestions for Improvement: This includes recommendations from learners for integrating technology, providing clearer guidance, and enhancing discussion-based learning to boost engagement and comprehension.

Step 6: Producing the Final Analysis

Finally, the researcher synthesised the above themes into a coherent narrative in a way that highlighted their relevance to the research question.

Merging of thematic codes. Initially during the analysis of the RQ4 dataset, the researcher examined students' responses from the PF and DI groups separately, as had been done with the previous two qualitative datasets (i.e., the responses to Challenge Problems 1 and 2, and the responses to the Compare-Contrast Task of Challenge Problem 3). This approach allowed a close examination of the unique ways students shared their learning experiences within each instructional context. However, as the analysis progressed, it became apparent that many concepts expressed by students in both groups had similar underlying meanings. These included thoughts on reasoning, engagement, difficulty, and learning transfer.

After receiving feedback from the two external qualitative coders, the researcher decided to merge the two initial codebooks (one for the PF data and the other for the DI data) into one unified framework. Both experts independently reviewed the coding and suggested that combining the codebooks would enhance clarity and allow for better comparisons between instructional groups. This integration helped the researcher to present a more effective interpretation of the similarities and differences between PF and DI learners and improved the overall clarity and consistency of the coding process.

As an example of how the merging of thematic codes proceeded, it was observed that students in the PF group often highlighted the benefits of hands-on learning. They shared how this approach helped them experiment with models and think deeply about system behaviours. Initially, then, the relevant theme for the PF group was “Exploratory Learning and Model Testing.” In partial contrast, students from the DI group talked about similar experiences in terms of trial-and-error and visual understanding, but they talked less about having control over their learning. The corresponding code for the DI group was initially phrased as “Trial-and-Error Learning and Parameter Use.” Following consultation with the external coders, both sets of responses were grouped under the unified subtheme “Exploratory Learning and Experimentation,” which represents how active interaction with models encouraged students in both groups to have a deeper understanding of the models. Similarly, other themes were also merged, keeping in mind the relevance to the objective of the research question. This consolidation helped create a better codebook with meaningful themes that align clearly with instructional design recommendations.

Final analysis. The coding and analysis process of this dataset (i.e., Self-Reports and Focus-Group Transcripts) ultimately resulted in the identification of four main themes: “Engagement and Positive Learning Experiences,” “Challenges and Barriers to Effective Learning,” “Comparison of Learning Methods,” and “Suggestions for Improvement.” A comprehensive summary of the established themes and sub-themes of this dataset is presented in Table 21.

Table 21

Thematic Coding of the Self-Reports and Focus-Group Transcripts

Name	Description (# References)	Examples
Theme 1. Engagement and Positive Learning Experiences: Focuses on how both models foster engagement, critical thinking, and positive learning outcomes (158)		
Clarity and Comprehensiveness of Content	Highlights how materials aid in conceptual clarity and knowledge retention (37)	"Actually, the material was smooth, easy to absorb and understand. the sequence of activities simplified the content well Enough. Applying the knowledge on the simulator improved the understanding even more." (DI Group) "The content and length of videos are perfect...the concepts where very clear and straight forward to understand." (PF Group)
Enhancing Understanding Through Visual Aids	Emphasises the benefits of interactive tools in enhancing learning experiences (85)	"I understand how the learning process go after I done the first two session. the model help a lot with the video." (PF Group) "Especially for us medical students, I noticed that, eventually, we like to see more animated videos and animated examples to comprehend and to understand whatever topics that could be given to us. So this is a very big deal and a very big step to take, especially I think that when I'm answering the other questions too." (DI Group)
Exploratory Learning and Experimentation	Discusses how hands-on learning through models enhances engagement and discovery (19)	"In situations like epidemics, you need to know the whole idea, and you have and you have to learn about about it as a whole. And this what, agent based modle grant us to do, dynamic and interactive learning, and also experiential learning, better than passive learning, is very beneficial in, at this stage." (PF Group) "Actually, I remembered some of them, like, the to the tipping point, how it suddenly changed from one thing to another and the emergence of the other things." (DI Group)
Flexibility in Manipulating Parameters	Highlights the benefits of adjusting parameters to explore different scenarios (14)	"The second thing, which I, really enjoy and learn from it is playing with the parameters, in ABM system when we, used, used it for different, different concepts in malaria, in COVID 19, and other concepts which were in the experiment where we played with the factors. For example, the immunity period, the infectious, the infected people, the ICU capacity, and others other things." (PF Group)

User-Friendly Website for Learning Support	Focuses on how an intuitive website improves accessibility and understanding (2)	"Tying the SIERS model to a real-world disease kinda helped put everything into perspective more, as to what each parameter meant." (DI Group) "The model simulation was very good way to demonstrate idea discussed in each model." (DI Group)
Theme 2. Challenges or Barriers to Effective Learning: Identifies the obstacles that hinder learning in both models, such as limitations, complexity, lack of guidance, and technical difficulties (72)		
Complexity and Confusion	Highlights challenges in understanding complex charts, models, and parameter interactions (29)	"Marketing trend chart was difficult to comprehend, there were many trend lines and i couldn't link them together,,, also i couldn't link them with the changes I made in the parameters and how each parameter affect the results." (PG Group) "The charts and graphs sometime difficult to understand. If there is a way to make it easier." (DI Group)
Lack of Instructor Guidance	Discusses the need for clearer explanations and instructor feedback to enhance learning (14)	"I think it was everything was in point, but some of the parameters was not clear. If you could give us the definition before that, would be a bit better for us." (PF Group) "It was very clear for the other examples and for the other concepts that you gave us, like, the tipping points because I couldn't understand from the beginning before you gave us the videos and the explanation about it, and then we applied it on the system." (DI Group)
Limited flexibility	Examines how excessive parameter flexibility can lead to unrealistic or confusing scenarios (8)	"I mean, we have many parameters that that that's nice to change the situation of the model as you want. But at the same time, it create non realistic situation." (PF Group) "Yes. I faced that too. It's easy to forget when there's too much to do. I mean, when there are many parameters, I can't play with all of them at once." (DI Group)
Technical Difficulties	Identifies issues like software crashes and slow responses that disrupt learning (8)	"The program crashed while i was answering the last quiz so i had to use the information from the videos." (PF Group) "Because those models was very slow to respond. Okay. Maybe we do software bugs or something like that." (PF Group)
Unclear Question Formats	Focuses on the difficulty of interpreting and responding to complex or vague questions (13)	"I don't really like to answer paragraph questions, my ideas get lost and cant really remember the things that i have to right." (PF Group) "So about the study, sometimes I really forgot the the main aim, and then you remind us about it." (DI Group)

Theme 3. Comparison of Learning Methods: Compares PF with traditional learning methods, highlighting differences in learning retention, the effectiveness of simulations, and the value of face-to-face discussions (33)		
Enhanced learning through ABM	Focuses on how computational models provide a time-efficient and engaging way to understand complex systems (14)	"Learning based computer models was very good and time saving. Also it is very useful on having much results at short time." (PF Group) "Traditional methods, like lectures and textbooks, usually present information in a linear and static way. You learn about the concepts, the factors involved, and maybe see some examples or case studies. However, with ABMs, I was able to engage with the material more actively." (DI Group)
Limited Flexibility in Traditional Learning	Discusses the shortcomings of traditional learning methods, especially in retaining complex knowledge (14)	"I mean, usually, in traditional way of teaching, I mean, like, in lecturing approach, you you might hear too lots of information within 2 hours. And after that, just to go home and just to do the exams and whatever you need to remember." (PF Group) "It took us very long to understand even the question so that we can answer it very quickly. So, it's always harder to apply the knowledge without, like, using, and it even take longer time to memorize things. I remember myself, at the beginning of my years in the medical school, it was very hard for me to, memorize or even to learn a lecture. It took me about 8 hours sometimes because to understanding the topic." (DI Group)
Practical Application of ABMs in Learning	Examines how Agent-Based Models (ABMs) help relate abstract concepts to real-world applications (5)	"As I mentioned, it also helped us to, relate this to our reality and to apply this concept in different issues, not only in the health and medical issues, but in different issues and levels in our life. So yeah. Even though it's a little bit complex, but I benefit a lot from it." (PF Group) "It was very helpful because I could see what was happening in the system, I can imagine, and even I can be involved in the system manipulation. That gives me a better understanding of the different relationships." (DI Group)
Theme 4. Suggestions for Improvement: Focuses on recommendations (124)		
Application of ABM in other fields	Focuses on extending the use of ABMs to diverse areas such as public health and healthcare management (18)	"I think this approach could be very beneficial for certain field in medical medicine, for example, in public health. For example, like epidemics and infections to this." (PF Group) "The use of Agent Based Models and systems thinking could be applied to other complex topics, such as understanding the spread of non-communicable diseases, healthcare system dynamics, or even patient behavior and outcomes in response to different treatment plans." (DI Group)
Improving Content Depth	Highlights the importance of enriching the learning material with comprehensive, detailed content for clarity (54)	"I think if we had a session with further details about how we are expected to work through the models, it would be much better experience. the short videos provided were useful but knowing further details would have improved our performance. the questions were not that much clear to me and they required long answers." (PF Group)

		"But after some time, it would be very easy to forget whatever I have learned there. So if it was taken, like, a step by step, like every single topic could be divided into separate ones, that will be much easier for us, not only to learn, but also to remember whatever is given." (DI Group)
Integrating Technology to Enhance Traditional Learning	Focuses on blending traditional lectures with technology to deepen understanding and provide visual support (6)	"As you just say, in traditional way, we get the, information by lectures that is given by lecture to to understand the way evidence emerged. But, here we use computer installation to better visualize and understand them... I just said, it's better to implement it in the course of epidemics to see if the student used it." (PF Group) "It is possible to use different models in order to facilitate understanding of how diseases spread, the different factors that affect their spread, and the different methods of their spread. It is also possible to use educational videos, additional information, and studies for various diseases." (DI Group)
Need for clearer guidance	Emphasises the need for more detailed instructions and explanations to help learners navigate each step effectively (19)	"I think if we had a session with further details about how we are expected to work through the models, it would be much better experience. the short videos provided were useful but knowing further details would have improved our performance. The questions were not that much clear to me and they required long answers." (PF Group) "The experience was completely asynchronous, and I think adding some synchronous meeting to explain how to use the model may be better although the asynchronous can achieve the needed goals." (DI Group)
Preference for Discussion-Based Learning	Highlights the value of discussions in reinforcing understanding (19)	"Being physically present in a class where the teacher can clarify some points if not understood." (PF Group) "We don't really practice, like, quizzes or and, like any assignments after afterwards, like, after the lecture. But we also have other courses where it's, more of a discussion, thing. Yeah. And we have other classes in other courses, which we call them tutorials and the lab classes, but we, more discuss than we take or absorb information. And for us, at least for me, like, personally, I prefer to discuss more with the doctor, which make me, you know, like, questioning myself and questioning my knowledge." (DI Group)
Simplifying Question Formats	Discusses the need for clearer and more accessible assessments (8)	"I don't really like to answer paragraph questions; my ideas get lost and can't really remember the things that I have to write." (PF Group) "way of asking and answering could be easier by MCQs." (DI Group)

Inter-rater reliability of focus-group interviews and self-report data coding. As with the previous qualitative datasets, typical inter-rater reliability measures (e.g., Cohen's Kappa) were not calculated for this dataset. Instead, the two external coders co-coded a randomly selected subset of 20% of the student responses, stratified by group condition. As before, the coders independently reviewed their coding, openly discussed any discrepancies, and agreed upon refinements to the codes for the sake of clarity and consistency. A summary of this refinement process for the focus-group and self-report data is provided in Table 22.

Table 22

Inter-Rater Reliability of Focus-Group Interviews and Self-Report Data Coding

Dataset Segment	Level of Agreement	Key Discrepancies	Resolution Approach	Outcome
Focus Group & Self-Report (20% subset) – Researcher vs. Coder A	High agreement with some interpretive variation	Differences in identifying "Exploratory Learning" vs. "Trial-and-Error" themes	Clarified hands-on interaction criteria; added examples to show active vs. passive engagement	Codebook refined to distinguish active experimentation from observational learning
Focus Group & Self-Report (20% subset) – Researcher vs. Coder B	Substantial agreement	Overlap in coding "Visual Aids" vs. "Content Clarity"	Reviewed student quotes; refined definitions to separate support tools from conceptual presentation	Codebook updated to maintain clear boundaries between visual supports and overall material clarity
Focus Group & Self-Report (20% subset) – Researcher vs. Both Coders	General alignment	Confusion over merging "Suggestions for More Guidance" with other improvement themes	Joint discussion on instructional needs vs. content depth; examples reviewed collaboratively	Codebook updated to ensure consistent application of "Clearer Guidance" and "Improving Content Depth" themes

Trustworthiness: Reliability and Validity

Trustworthiness is vital to the research process and arises when research is conducted rigorously and the findings are deemed reliable, valid, and applicable (Enworo, 2023). Broadly speaking, trustworthiness is the extent to which the findings of a study are dependable, credible, and transferable. *Dependability* refers to consistency and stability of results over time and various conditions (Enworo, 2023; Lincoln & Guba, 1985). It is achieved in the current study through the maintenance of an audit trail by documenting in detail the raw data collection procedures, steps of analysis, and other major decisions made that affected the research design. The study's methodological rigour has been enhanced by the detailed descriptions of any changes made during the study.

Credibility can be defined as the extent to which findings of the study accurately represent the participants' data and experiences (Creswell & Poth, 2016). This study has achieved credibility by relying on multiple methodologies (quantitative and qualitative) and data sources (pretest and post-test online written responses, sessional challenge problems, focus group interviews, and self-reports). Previous research suggests that multiple methodologies and data sources can significantly enhance a study's credibility, as different methods help cross-verify the results, thus increasing validity (Hussein, 2009). Finally, *transferability*, that is, the extent to which results of the study can be replicated in other settings or contexts (Lincoln & Guba, 1985), is also achieved in this study by the provision of detailed descriptions of the processes, participants, and context of the study so as to allow readers to judge their relevance to the study's findings.

Overall, these three components, dependability, credibility and transferability have been applied during the collection of both data types, experimental and qualitative (focus group and self-reports) to increase the reliability and validity of the collected data. In any research study, addressing potential threats to reliability and validity is crucial to ensure the accuracy and credibility of the findings. This study carefully considered various factors that could affect internal validity and implemented several strategies to mitigate these risks. The following sub-sections explain these strategies and how they were applied to each type of research approach separately.

Reliability and Validity (Trustworthiness) of Experimental Data

There were quite a few potential issues identified with the collection of experimental data which would decrease the reliability and validity of the results. Some issues were resolved before the data collection was started and others were addressed using different strategies during the data-collection period. For example, before the data collection started, the researcher piloted the novel ABMs to check their functionality before including them in the main study. Feedback from users validated the model's overall design and helped the researcher identify certain functional issues, such as inconsistences or errors in the model, thus improving the models' usability and level of engagement. Moreover, the scenarios and parameters for each ABM were checked by two modelling consultants and two epidemiologists from different universities. This interdisciplinary rigour safeguarded the scientific accuracy and relevance of the session content, strengthening its usability and improving the overall credibility of the study.

The same subject-matter experts (epidemiologists) examined each session's content, including the content scripts used to record the content videos, along with all test questions, for accuracy and alignment with the study's goals. Validation of the content in this way not only enhanced credibility and reduced biases or inaccuracies in the material but also improved the clarity and relevance of the content and the test questions. It is important to mention that the structure of the tests, the rubric, and the sequence of the instructions for both groups were adapted from Jacobson et al. (2020, 2017). The use of an established framework in this way enhances both reliability and content validity. Finally, to improve the reliability of the study, the inter-rater reliability of the two scorers was computed. It was found, as noted earlier, that high degrees of reliability were achieved. Moreover, the raters were drawn from the medical field to avoid undue subjectivity and increase the reliability of the scores.

As noted in the introductory chapter, epidemiology shares several core characteristics with other complex systems, such as emergent behaviours, nonlinear interactions, and adaptive components. However, epidemiology presents unique challenges because it involves a range of behavioural, social, and biological factors (Turnbull et al., 2018; Montefusco & Angeli, 2024). Like climate systems and natural ecosystems, epidemiological models encompass path dependencies and feedback loops (Mudd et al., 2024); however, unlike physical (for example, traffic patterns) or abstract systems (for example, stock markets), epidemiological models must consider ethical constraints and human agency (Turnbull et al., 2018; Montefusco & Angeli, 2024). In complex systems, previous work on PF suggests two main boundary conditions. The first is that the effectiveness of PF diminishes when there are limited foundational domain-specific schemas among learners to structure their failed attempts (Nachtigall et al., 2020; Steinhorst et al., 2023). The second condition is that the benefits associated with PF depend on subsequent instruction, linking emergent patterns to basic mechanisms (Steinhorst et al., 2023; Montefusco & Angeli, 2024).

Extensive interdisciplinary discussions with specialists during the study's design phase guaranteed that the concepts of complex systems selected in the study (for example, emergence and dynamic equilibrium) were rigorously mapped to epidemiology principles. This helped address gaps in previous PF-based studies for non-STEM domains. The concepts of complex systems are broadly transferable; however, their pedagogical application needs adaptations that are field-specific (Mudd et al., 2024). For example, the present study structured challenge problems to compare ABMs of epidemics (disease spread) with forest fires, an approach that fostered analogical reasoning and reduced a boundary condition of PF, namely, insufficient foundational schemas. The study also mitigated the above conditions via scaffolded reflection exercises, which linked initial model failures of students with fundamental epidemiological principles. The study design also emphasised sequencing instruction, emphasising the link between population-level emergent outcomes and micro-level transmission mechanisms (Montefusco & Angeli, 2024).

During the data collection period, there was no communication between students in the experimental and control groups. However, even though it was possible that students might discuss the study with each other enough to recognise differences in the instructional approaches to which each was exposed, it was deemed unlikely that this awareness alone would have significantly altered the problem-solving strategies they employed. Other issues included the medium of data collection for experimental data which was online. As the data collection instrument was posted online on a learning management system (Moodle) as explained earlier, there were several concerns regarding the truthfulness in the responses of the students. Here, the study adopted blinding strategy, whereby students were placed in the experimental or control group without telling them which group they are part of. In this manner the students were less likely to be aware of the distinctions between groups, decreasing biases in their responses and behaviour. One other strategy that was adopted included reinforcement of data anonymity. Students were informed that their responses used in the research will not be directed towards their personal identity which mitigated the risk of students being conscious of their answers and using other different misleading ways of responding to the questions.

Furthermore, given the risk of online data collection and whether the models within the experiment were effective enough for the students, the study also implemented triangulation of data strategy. This means supplementing online data collection with other methods like interviews or focus groups to extract feedback from the students and cross-verify the findings. The current study used focus groups, which added multiple data sources to the current research methodology and enhanced the study’s validity by confirming the findings.

While all these strategies were considered and implemented prior to and during the data collection process, it should be noted there were several issues of low responses ratings and delayed submissions. However, additional follow-ups in the analysis phase—including the use of statistical tests such as Cronbach Alpha—were carried out to ensure that these issues did not impact the credibility, transferability and dependency of the results or the collected data.

Reliability and Validity (Trustworthiness) of Qualitative Data

For qualitative data collected via focus groups, some potential issues that were faced in pre-, during- and post-data collection stage were also addressed in the best of the researcher's ability through several strategies. For example, in pre-data collection stage, there were concerns regarding whether the participants would accurately represent the existing proportion of the students which might lead to skewed data. However, the students' representation was ensured to be proportionate to the experimental and control groups, as established beforehand. Moreover, both male and female representation was also ensured, which enable gather diverse viewpoints.

During the data collection stage, the concerns mostly revolved around the adequate flow of the discussion. Even though the data collection via focus groups were taking place on online platform, there was a problem of dominance by specific individuals. As discussed earlier in the chapter, this issue was also pointed out by multiple female participants that they would like to partake in discussions where the ratio of male participants to female is less so they can voice their opinions without the fear of judgement or any hesitation. To address this issue, facilitation strategies were used, such as dividing participants of the focus groups based on gender (separate male and female groups) to mediate and encourage each participant to speak easily and involving discussion strategies of answering turn-by-turn. Additionally, there were also issues relating to irrelevant discussion flow (Karunarathna et al., 2024). To guarantee that the focus group discussion is relevant, and all the information being generated helps the current research to reach conclusions for the research questions, the researcher designed focus group interview-guide. The predetermined questions in the guide helped the moderator to maintain the relevancy of the discussion.

Finally, post-data collection issues compelled the researcher to adopt the Excel spreadsheet for analysis as information collected through focus group is difficult to analyse and extract patterns from. It simplifies coding qualitative data and identifying patterns based on common keywords and phrases used by respondents. The tool leads to relatively less chances of error ensuring that the results are reliable and valid.

It should be noted, that despite the strategies implemented and issues accounted for, multiple participants failed to show up on the zoom call making the amount of information that was collected less than expected. However, such last-minute problems are out of control and it was ensured that regardless of the participants who were not available the study had enough data to draw conclusions from.

Chapter Four: Results

This chapter presents the results of the study outlined in the preceding chapter, organised here according to the research questions (RQs) presented earlier. The broad purpose of this analysis is to determine the effectiveness within an undergraduate epidemiology pedagogical setting of a Productive Failure (PF) instructional design versus a Direct Instruction (DI) design for promoting the learning of declarative and explanatory knowledge of epidemics and complex systems concepts and for supporting the transfer of such concepts to solve new problems in epidemiology and other domains.

RQ1: No Advantage of PF for Declarative or Explanatory Knowledge

The first research question explored whether students in the PF or DI learning condition would exhibit better learning outcomes across three types of instructed knowledge: (1) declarative knowledge of epidemics, (2) declarative knowledge of complex systems concepts in epidemiology, and (3) explanatory knowledge of complex systems concepts in epidemiology. For those dependent variables that were measured by multiple questions (e.g., questions 1-4 on the pretest and post-test, which assessed the type (1) knowledge above), responses were averaged to create overall scores. Table 21 presents a representative sample of the pretest and post-test questions associated with each dependent variable related to RQ1 (cf. Table 5 in Chapter Three).

Table 23

Pretest/Post-test Questions and Corresponding Dependent Variables for RQ1

Dependent Variables	Questions
Declarative knowledge of epidemics concepts	1. What are the modes of transmission of infectious organisms? 2. What does mortality refer to in epidemics? 3. What is the difference between latency and incubation period? 4. Please describe what disease prevalence is.
Declarative knowledge of complex systems in epidemiology	5a. What are examples of dynamic equilibrium in epidemics? 6a. What are examples of tipping points in epidemics? 7a. What are examples of emergent properties in epidemics?
Explanatory knowledge of complex systems in epidemiology	5b. Please explain 6b. Please explain 7b. Please explain.

Analysis for RQ1

Before analysing the post-test scores for the dependent variables, an independent samples *t*-test was conducted on the pretest scores to ensure no significant group differences existed at baseline between the experimental (PF) and control (DI) conditions. The results indicated no significant differences between the groups across all dependent variables: declarative knowledge of epidemics ($t(33) = 0.711$; $p = 0.483$), declarative knowledge of complex systems in epidemiology ($t(33) = -0.401$; $p = 0.691$), and explanatory knowledge of complex systems in epidemiology ($t(33) = 0.029$; $p = 0.977$). These findings confirm that the groups were comparable prior to the intervention, allowing greater confidence that any observed differences at post-test are likely attributable to the instructional methods received rather than preexisting disparities.

According to the Bayesian Independent Samples t-tests performed, moderate evidence was observed for H_0 (no difference) across the three dependent variables (all pre-test measures): declarative knowledge (epidemics) ($BF_{10} = 0.399$, error < 0.01%), explanatory knowledge (complex systems) ($BF_{10} = 0.328$, error < 0.01%), and declarative knowledge (complex systems) ($BF_{10} = 0.349$, error < 0.01%). These results show that prior to the intervention the PF and DI groups were equivalent, thus strengthening confidence that any observed post-intervention effects were not attributable to disparities at the baseline (see Appendix L).

The results of the main analyses comparing post-test scores across conditions are presented in detail below, organised by dependent variable.

Declarative Knowledge of Epidemics. An independent samples *t*-test was conducted to determine whether students in the PF or DI conditions demonstrated superior learning outcomes in declarative knowledge of epidemics. The results indicated no statistically significant difference between the two groups: $t(33) = 0.750$, $p = 0.458$, $d = 0.256$.

Specifically, students in the PF condition (*M* = 2.18, *SD* = 0.64) did not differ significantly from those in the DI condition (*M* = 2.00, *SD* = 0.73) in terms of declarative knowledge of epidemics exhibited on the post-test. The effect size is also small as per Cohen's (1988) thresholds.

Declarative Knowledge of Complex Systems in Epidemiology. An additional independent samples *t*-test examined potential differences between the PF and DI conditions in their declarative knowledge of complex systems in epidemiology. Although students in the DI condition (*M* = 1.51, *SD* = 1.07) scored somewhat higher than those in the PF condition (*M* = 1.28, *SD* = 1.06), the analysis again revealed no statistically significant difference between the two groups: $t(33) = -0.63$, $p = 0.536$, $d = -.213$.

Explanatory Knowledge of Complex Systems in Epidemiology. Finally, an independent samples *t*-test compared students in the PF and DI conditions on explanatory knowledge of complex systems in epidemiology. Similar to the other outcomes, the analysis found no statistically significant difference between the two groups: $t(33) = -0.07$, $p = 0.944$, $d = -.024$. That is, students in the PF condition (*M* = 1.08, *SD* = 1.15) and those in the DI condition (*M* = 1.11, *SD* = 1.14) demonstrated comparable explanatory knowledge.

The Bayesian Independent Sample t-tests exhibited moderate evidence supporting H_0 across the three post-test dependent variables: declarative knowledge (epidemics) (BF_{10} = 0.408, error < 0.01%), explanatory knowledge (complex systems) (BF_{10} = 0.328, error < 0.01%), and declarative knowledge (complex systems) (BF_{10} = 0.381, error < 0.01%). The Bayes factor, when converted ($BF_{01} = 1/BF_{10}$), showed that the data were more likely (2.44 to 3.03 times) under the H_0 than under H_1. Thus, the post-test Bayesian analysis revealed anecdotal to moderate evidence for H_0, suggesting no meaningful group differences (see Appendix L).

Pre-Post Comparison. To examine whether the students in the PD or DI learning condition exhibit better learning outcomes across the three types of instructed knowledge, paired-sample t-tests were also conducted. The tests examine significant different differences between pre- and post-test scores; it was done for both experimental and control groups.

Within the PF group, when declarative epidemics scores were assessed, significantly higher post-test scores were observed ($M = 2.18$, $SD = 0.64$) when compared to pre-test scores ($M = 1.63$, $SD = 0.62$), $t\ (19) = -4.10$, $p < 0.01$. Cohen's d is equal to -0.92, suggesting a large negative effect size. In the case of declarative complex scores, significantly higher post-test scores were observed ($M = 1.28$, $SD = 1.07$) when compared to the pre-test scores ($M = 0.75$, $SD = 0.76$), $t\ (19) = -3.14$, $p < 0.01$. The effect size can be termed as medium to large ($d = -0.70$). When explanatory complex systems' scores were assessed, again the post-test scores were significantly higher ($M = 1.08$, $SD = 1.15$) than the pre-rest scores ($M = 0.25$, $SD = 0.66$), $t\ (19) = -3.08$, $p < 0.01$. The effect size was termed as medium to large ($d = -0.69$).

Within the DI group, when declarative epidemics scores were examined, the post-test scores were found to be significantly higher ($M = 2$, $SD = 0.73$) than the pre-test scores ($M = 1.47$, $SD = 0.70$), $t\ (14) = -2.45$, $p < 0.05$. The effect size was medium ($d = -0.63$). When declarative complex scores were assessed, significantly higher post-test scores were observed ($M = 1.51$, $SD = 1.07$) when compared to the pre-test scores ($M = 0.84$, $SD = 0.59$), $t\ (14) = -3.57$, $p < 0.01$. The effect size can be defined as large ($d = -0.92$). Lastly, in the case of explanatory complex systems' scores, the post-test scores were significantly higher ($M = 1.11$, $SD = 1.14$) than the pre-test scores ($M = 0.24$, SD = 0.39), $t\ (14) = -3.26$, $p < 0.01$, with a large effect size ($d = -0.84$). The results exhibit that both PF and DI groups showed significant improvements across the given constructs; however, it does not help prove differences between PF and DI conditions.

Strong evidence for improvements was shown by the Bayesian-based paired samples t-tests across the three measures, pre-test to post-test within the PF group: epidemics declarative knowledge ($BF_{10} = 56.764$, error $< 0.001\%$), complex systems explanatory knowledge (BF_{10} = 7.629, error < 0.001%), and complex systems declarative knowledge (BF_{10} = 8.571, error < 0.001%). In declarative epidemics knowledge, very strong evidence was observed in declarative epidemics knowledge based on the Bayes factor (BF_{10} > 30), while for improvements in explanatory and declarative complex systems knowledge, strong evidence was indicated (BF_{10} > 6).

Within the DI group, mixed evidence for pre-post improvements was found. In the case of epidemics declarative knowledge, there was limited evidence for improvement (BF_{10} = 2.404, error < 0.001%). On the other hand, strong evidence for improvement was demonstrated with regard to complex systems declarative knowledge (BF_{10} = 14.814, error < 0.01%) and complex systems explanatory knowledge (BF_{10} = 8.868, error < 0.001%) (see Appendix L).

Summary of Declarative/Explanatory Knowledge Results (RQ1)

In summary, the statistical analyses did not identify significant differences in learning outcomes on declarative or explanatory knowledge between the PF and DI conditions; moreover, the effect sizes reported were small, suggesting no practical difference between the groups. Regarding RQ1 and its associated hypotheses, the lack of a significant difference between the performances of the two groups on both types of declarative knowledge (i.e., of epidemics concepts and of complex systems concepts in epidemiology) means that nondirectional H1 was confirmed:

> **H1** (nondirectional): The PF-treatment group will not exhibit better learning outcomes than the DI-control group on *declarative* knowledge of epidemics or *declarative* knowledge of complex systems in epidemiology.

However, the lack of a significant difference in the performances of the two groups on explanatory knowledge of complex systems concepts in epidemiology means that H2 was *not* confirmed:

> **H2** (directional): The PF-treatment group will exhibit better learning outcomes than

the DI-control group on *explanatory* knowledge of complex systems in epidemiology. This means that the present study was unable to corroborate in the field of epidemiology pedagogy the previous finding (Jacobson et al., 2017) that PF offers a significant advantage over DI when it comes to the learning of explanatory knowledge.

RQ2: PF Provides Advantage for Near Within and Far Across Domain Transfer

RQ2 aimed to examine whether students in the PF treatment condition, compared to those in the DI control condition, would demonstrate an enhanced ability to transfer their knowledge of epidemics concepts and complex systems concepts in epidemiology to new content. This transfer was assessed both within the domain addressed during instruction (near within domain transfer) and to a domain not explicitly covered in the study (far across domain transfer). Table 24 presents a representative sample of the pretest and post-test questions associated with each dependent variable related to RQ2 (cf. Table 5 in Chapter Three).

Table 24

Pretest/Post-test Questions and Corresponding Dependent Variable for RQ2

Dependent Variables	Questions
Near within domain transfer problems	*"Common Cold" problem* Key questions: (a) Discuss the disease and its ability to move through the population. (b) With the information that you have regarding this disease, what preventative measures can be put into place? *"Salmonella Enterica" problem* Key question: What are the differences between an epidemic and an outbreak of a disease in relation to the different population sectors involved in the proliferation of the disease?
Far across domain transfer problems (post-test only)	*"Road Injury and Mortality" problem* Key questions: (a) Can you explain the relationship between mobile phones and road injury and mortality and how prevention relates to data surveillance? (b) Please describe and explain why different preventative and regulatory measures are used for different members of the population to reduce the rate of injury and mortality associated with mobile phones?

Analysis for RQ2

A preliminary analysis of the two groups' pretest near within domain transfer scores (similar to that performed for RQ1) was conducted to determine whether any significant difference between the groups existed at the baseline. The results showed no significant difference ($t(33) = 1.287$, $p = 0.207$), confirming that the groups were comparable prior to the intervention. The baseline equivalence was also examined by a Bayesian independent sample t-test on the near transfer (pre-test) score, yielding $BF_{10} = 0.619$ (error < 0.01%). This result showed that the data were 1.61 times (1/0.619) more likely under H_0 than H_1, providing anecdotal evidence for H_0 as per Jeffreys' (1961) evidence categories (see Appendix L).

The following tests include within-group (for both PF and DI) and between groups (PF vs. DI) analyses.

Near Within Domain Transfer. For the PF treatment group, a paired samples *t*-test was conducted to evaluate the effect of the PF instructional/learning approach on students' ability to transfer knowledge within the instructed domain (near within domain transfer). The results revealed a significant improvement in near within domain transfer scores following the PF intervention. Specifically, students' near -within domain transfer scores increased from the pretest ($M = 1.35$, $SD = 0.59$) to the post-test ($M = 2.00$, $SD = 0.82$), with this difference being statistically significant ($t(19) = -4.33$, $p < .001$, $d = 0.67$). The effect size exceeded Cohen's (1988) convention for a medium effect ($d = 0.50$). These results suggest that the PF approach is effective at enhancing students' ability to transfer knowledge already learned to new problems within previously instructed contexts.

In near transfer ability (pre- to post-test), very strong evidence for substantial improvement as a result of the PF intervention (per Jeffreys' 1961 criteria) was revealed by the Bayesian paired-samples t-test ($BF_{10} = 89.856$, error < 0.001%).

For the DI control group, the results indicated no statistically significant improvement in near within domain transfer scores from the pretest (*M* = 1.06, *SD* = 0.70) to the post-test (*M* = 1.46, *SD* = 1.07): $t(14) = -1.31$, $p = .21$, $d = 0.34$. Although the effect size ($d = 0.34$) falls within the range of a small effect, it does not meet Cohen's (1988) threshold for a medium effect ($d = 0.50$), suggesting that the DI approach may not be effective in fostering students' ability to apply previously learned knowledge to the solution of new problems in the same, previously instructed context.

The Bayesian paired-samples t-test for the DI group, comparing near-transfer pre- and post-test, yielded $BF_{10} = 0.538$ (error = 0.018). For H_0, this showed weak evidence according to Jeffreys' (1961) scale (see Appendix L).

To examine whether the PF treatment group outperformed the DI control group (between-group analysis) in the near transfer scores, an independent samples *t*-test was performed on both groups' post-test near within domain transfer scores. The results are not significant: $t\ (33) = 1.67$, $p < 0.1$, $d = 0.570$. However, the effect size suggests a medium effect, showing a meaningful practical difference between the conditions. In addition, the students in the PF condition (*M* = 2.00, *SD* = 0.81) demonstrated substantially higher near within domain transfer scores compared to students in the DI condition (*M* = 1.47, *SD* = 1.07). Given the nonsignificant result, it cannot be confidently stated that the PF approach was any more effective in enhancing students' ability to apply learned concepts to closely related problems within the same domain than was the DI approach.

A Bayesian independent sample t-test was run on the near transfer post test scores, comparing the PF and DI groups, yielding $BF_{10} = 0.947$ (error < 0.01%). Based on the inverse Bayes factor, which was equal to $BF_{01} = 1.06$, the result fell under the range of anecdotal evidence, suggesting insufficient data to meaningfully differentiate between H_0 and H_1.

Far Across Domain Transfer. To evaluate whether the PF treatment group outperformed the DI control group in applying knowledge of epidemics concepts and complex systems concepts to a new, uninstructed domain (i.e., far across domain transfer), an independent samples *t*-test was conducted on both groups' post-test far across domain transfer scores.

The analysis revealed a statistically significant difference between the groups: $t(33) = 4.17$, $p < .001$, $d = 0.65$. Moreover, the effect size exceeded Cohen's (1988) convention for a medium effect ($d = 0.50$), indicating a meaningful practical difference between the conditions. Specifically, students in the PF condition ($M = 2.60$, $SD = 0.59$) demonstrated significantly higher far across domain transfer scores compared to students in the DI condition ($M = 1.66$, $SD = 0.72$). These findings suggest that the PF approach is more effective than the DI approach in engaging learners to transfer their learned knowledge of epidemics and complex systems to tackle unfamiliar problems in different domains.

Decisive evidence for group differences was revealed by the Bayesian independent samples t-test in far transfer post-test scores ($BF_{10} = 116.012$, error < 0.001%), considering the classification criteria of Jeffreys (1961), where $BF_{10} > 100$ indicates decisive evidence. This result showed substantial group differences in far transfer ability, supporting H_1.

Summary of Transfer-Related Results (RQ2)

In regard to RQ2 and its associated hypotheses, the above results mean that both directional transfer-related hypotheses, H3 and H4, were confirmed by the data:

H3 (directional): The PF-treatment group will exhibit greater gains than the DI-control group on *near within* domain transfer problems.

H4 (directional): The PF-treatment group will exhibit greater gains than the DI-control group on *far across* domain transfer problems.

That is, the present study replicated with epidemiology students the findings of previous research in other fields that a PF instructional design offers advantages over a DI design when it comes to promoting students' ability to transfer previously learned knowledge for use in solving new problems, particularly problems in domains outside the domain previously instructed (Jacobson et al., 2020; Kapur, 2008; Loewenstein et al., 1999).

RQ3: DI Group Shows Greater Terminological Precision; PF Group Shows Greater Awareness of Deeper, Structural Features of Models

RQ3 focused on the session-by-session learning process, specifically, whether and how the contrasting sequences of ABM-based problem-solving tasks in the PF versus DI designs might affect the learning process across multiple sessions differently. To answer this question, both quantitative and qualitative analyses were used: qualitative for challenge problems 1 and 2 as well as for the compare-contrast task of challenge problem 3, and quantitative for the application task of challenge problem 3 (refer to Chapter Three for details of these various tasks).

Analysis of Challenge Problems 1 and 2 Data

Types of Ideas Generated (Theme 1). This section presents the analysis of the types/diversity of ideas generated by students from both groups, i.e. experimental and control in Session 1 of the study. The study explored the use of Agent-Based Models (ABMs) by

Productive Failure (PF) and Direct Instruction (DI) groups to examine epidemic dynamics and emergent behaviours. The analysis uses a coding framework based on Kapur's methodology from 2008 and incorporates both primary themes and sub-themes from the qualitative codebook. Students used the SEIRS epidemic model to determine which factors influence disease spread and persistence within populations as part of their challenge problem. The ideas that emerged during the analysis are divided into the following categories (See Appendix J):

1. Biological & Epidemiological Factors – Mutation, immunity loss, transmission mechanisms.
2. Human Behaviour and Social Factors – Quarantine, hygiene, policy compliance, misinformation.
3. Mathematical and Simulation-Based Insights – Feedback loops, equilibrium, system behaviours.
4. Public Health and Policy Considerations – Healthcare infrastructure, government interventions.

Epidemic Dynamics and Emergent Behaviour (Session 1) - Challenge Problem 1: SEIRS Model – Epidemic Spread. Students worked on the first challenge problem which involved exploring the SEIRS epidemic model to determine the factors that influence disease spread and persistence in populations. Students who participated in the PF condition produced numerous ideas that spanned multiple sub-themes.

Biological and Epidemiological Factors

Students pinpointed essential epidemiological processes that influence how diseases spread, "That needs to be exposed to the infection… also their immunity against the disease should be low." Some students recognised reinfection cycles, linking them to immunity loss:

"After a high percentage of the population gets infected, they will have immunity for some time. After this time, immunity will be low again, and then we will have more cycles of the infection." Others mentioned asymptomatic carriers and latency periods: "The infection might be carried in their body for a period of time until the symptoms appear."

Human Behaviour and Social Factors

Several students discussed the role of human interactions in spreading disease: "The spread of infection is because of the connection between peoples every day as they communicate together." Students also highlighted the important role of quarantine in controlling disease transmission: "If we have an infected person, we need him to go in quarantine to have less connection."

Some of the students mentioned sociocultural factors, such as population density and hygiene influencing the spread of the disease. According to them, people crowded on each other like in China and India without policies will increase problem. A respondent says, "There are no ways to protect susceptible people from getting infected with the disease."

Mathematical and Simulation-Based Insights

The findings indicate that though some students attempted to describe systemic relationships in the model, their reasoning was often exploratory rather than precise. According to a student, "When the number of people contacted per day decreases, the immune and infectious increase while susceptible people decrease." Moreover, another one says, "The more the infection period is, the less the number of immune people, and more susceptible people."

Challenge Problem 2: Forest Fire Model: Emergent Behaviour. In the next session, students were asked to work with a forest fire ABM to explore emergence and micro-to-macro interactions in a system. The findings indicate that students from the PF groups generated many different analogies between fire spread and epidemic transmission. All the views shared by the students during this session are divided into sub themes which are mentioned below.

Fire Spread Analogies

Students highlighted the connections between fire spread patterns and disease transmission by showing how trees operate as fire hosts in the same way that humans serve as infection hosts. This approach attempts to use epidemiological principles to investigate physical phenomena. One student says, "The fire is like the infection, and the trees in the model are like the people in the community."

Mathematical and Simulation-Based Insights

Some students equated fire spread to disease spread, drawing direct comparisons. Students began to appreciate system dynamics when they realised that fire spreads through state changes in individual trees without physical movement. This demonstrates students' preliminary understanding of how complex systems produce emergent features. Many students recognised emergent behaviour and attempted to describe it: "The fire is not moving from place to another but spreading in different directions… it can spread to other trees, which changes and show us like moving fire."

Environmental and External Influences

Some students identified environmental influences on fire spread such as wind direction as a crucial effect in fire spread, demonstrating an increased awareness that external factors influence system behaviour. For instance, one student says, "The direction of the winds affects the fire movement."

However, on the other hand, students from DI group provided fewer but factually correct explanations of fire behaviour, "The movement is the emergent behaviour that simulates many trees burning… it looks like the fire is moving but that is only the fire spread from one tree to another." The findings indicate that DI students presented a direct cause-and-effect analysis of fire spread while the PF group created conceptual links between fire propagation and disease transmission dynamics. DI students provided mechanistic explanations that did not demonstrate the cross-domain reasoning abilities which PF group members showed.

In conclusion, during Session 1, PF students generated more ideas through cross-disciplinary exploration than DI students who provided structured yet limited responses. The study results validate the Productive Failure hypothesis which claims that early exploration leads to varied but disorganised thought processes whereas direct teaching results in distinct but limited conceptual knowledge.

Disease Transmission and Consumer Behaviour (Session 2)- Challenge Problem 1:

Malaria Model – Disease Transmission. This section explores the types of ideas produced by both PF and DI groups in response to the Malaria Transmission ABM challenge problem.

Based on responses, this theme is divided into three sub-themes:

1. Disease Transmission Mechanisms: Explains how malaria spreads.
2. Environmental and Population Factors: Explains how location, mosquitoes, and human behaviour influence transmission.
3. Prevention and Control Strategies: Discusses ideas about stopping malaria spread.

Disease Transmission Mechanisms

The findings of the study indicated that most of the students correctly identified that malaria is transmitted by mosquitoes carrying the Plasmodium parasite and described the infection process in detail. Some of their responses include, "Malaria is mosquito-borne infectious disease caused by the parasite." And "When mosquito with the parasite bites person, the parasite enters their blood and cause malaria infection."

When discussing various factors that influence the relationship between mosquito population levels and infection rates, the participants showed a complete understanding of vector-dependent transmission. Some participants discussed how rising mosquito populations increase disease transmission risks and also talked about how environmental factors such as temperature, humidity, and nesting locations influence this process. According to a respondent, "When the number of mosquitoes increase, there will be a high increase in infections and deaths in the hospital, ICU, or outside."

Environmental and Population Factors

Many students acknowledge that environmental factors play a crucial role in malaria transmission, particularly recognising mosquito breeding sites and human exposure patterns as major factors. Furthermore, the students identified vectors as essential components affecting both how diseases spread and where transmission occurs. One of them says, "Malaria spreads in a specific area either there is a vector that spreads it or expose to if not vector."

Some students examined the impact of human activity together with location-specific risks to evaluate disease spread across various exposure scenarios. One of their references is “The disease spreads when mosquito bites person at home or hospital or outside home. It could be also through infected blood.”

In conclusion, PF students successfully identified multiple environmental and behavioural factors that influence disease spread but lacked structured explanations connecting them into a broader system.

Prevention and Control Strategies

The findings also show that students also suggested prevention and control strategies, showing an awareness of both personal and environmental control methods. One of the quotes states, "People can use different measures to decrease getting infected, such as repellents, nets, and clothes."

While discussing the strategies, few responses recognised long-term strategies, such as biological control using natural predators. For example, "If we use predators in the environment like insects will decrease the number of mosquitoes and people getting infected." Furthermore, several students demonstrated a clear understanding that malaria cannot be directly transmitted from person to person, highlighting the significance of controlling mosquitoes rather than isolating afflicted individuals: "Malaria is not disease that spreads from person to person, so the way to protect people is to prevent mosquito bites."

PF students demonstrated creativity and engagement through their numerous ideas, yet their explanations showed structural weaknesses especially when explaining interactions between different factors. Participants centred their answers on singular incidents rather than population-wide mechanisms which show that their basic level comprehension has not yet embraced broad system thinking.

Challenge Problem 2 (Marketing Model – Consumer Behaviour). This section examines the types of ideas generated by both the PF and DI groups in response to the Marketing ABM challenge.

Based on their solutions to the problem given, this theme is divided into three sub-themes:

1. Consumer Decision-Making & Marketing Strategies
2. Psychological & Social Influences
3. Tipping Points & Product Longevity

Consumer Decision-Making and Marketing Strategies

The findings of the study show that in the Marketing ABM challenge, students examined the impact consumer decisions have on product adoption rates. PF students examined different elements that drive purchase choices while focusing on social behaviours:

"People's choices can be affected by advertisement on social media, someone telling about or finding product by chance."

Some students highlighted the connection between economic status and products while discussing the customers' buying behaviour. According to a PF group student, "People might choose products because of their education and income. Also, if we Provide good customer service, it will build trust."

DI students, on the other hand, discussed marketing concepts like brand loyalty and awareness and gave more structured responses: "Marketing influence tipping points. If product becomes popular, it can spread rapidly through advertising." Moreover, another DI student reports, "Social media trends and advertisements strongly impact consumer behaviour, even if there are better alternative."

PF students explored diverse topics including social behaviours along with economic factors and government policies to study disease distribution and marketing trends through a comprehensive yet occasionally disorganised method. The students often combined multiple ideas in their responses, but their inclusion of irrelevant elements made it difficult to discern clear cause-and-effect relationships, while students from the DI group presented structured and comprehensive explanations which utilised proven marketing tactics and public health methodologies. The PF students used exploratory thinking while DI students remained focused on strict analytical methods.

Psychological and Social Influences

The influence of psychological and social influences on consumer behaviour was another important sub-theme that emerged after the analysis of the students' responses. It provides views regarding how personal experiences, social norms, and peer interactions had a role in influencing participants' reasoning about why individuals adopt or reject products in the marketing model.

In the PF group, students frequently referenced individual-level behaviour and emotional factors, although their ideas were often exploratory and loosely structured. For example, one student noted: “People's experiences with products might be different from one person to another based on their educational and economic levels.”

Another student emphasised the influence of social learning and peer behaviour: “The experience of product can affect the people choices so Try the thing first and then decide.”

Students also mentioned social media and peer trends as influential. According to one of the students, “Social media and trends can affect people’s choices… people can be influenced by what others say about the product.” In contrast, students in the DI group provided more structured explanations that connected marketing success to social awareness and consumer psychology, for instance, “People’s choices are influenced by the rate of being aware of the product either by a friend, advertisement, or both.”

Another student reflected on how quality and word-of-mouth interact: “If the quality was good enough, people start to tell others about the product, so there will be more sales.”

These findings suggest that PF students explored a broader range of social and psychological variables, while DI students tended to give more targeted and well-structured responses, reflecting a clearer understanding of how psychological influences relate to consumer behaviour in the model.

Tipping Points and Product Longevity

The findings of the study indicate that the students from the PF group encountered difficulty in explaining the concept of the tipping point in marketing. The utilisation of metaphors that are not pertinent to the subject matter is evident in certain responses. These attempts demonstrate that the students initially encountered difficulty in describing systemic changes in consumer behaviour prior to receiving instruction. One of the students stated:

“It is useful for population to forget about your product, because if they see it again, they may change their mind.” Another participant shared, “The never buy group can’t change their mind. They are like a stubborn little cookie.”

However, the significant difference was observed in the DI group as their responses were clearer and better aligned with the concept of a marketing tipping point. As the students from the DI group had already received instructional guidance, they were able to provide structured explanations of how consumer choices evolve. According to one of them, “If a big number of mosquitoes become infected with malaria, the spread of the disease may reach a tipping point… Similarly, in marketing, small changes in customer choice can rapidly increase sales.” Another participant shared: “The tipping point in marketing is the moment when a product is highly accepted from consumers.”

These results indicated that the PF group generated exploratory but frequently ambiguous ideas, whereas the DI group produced more structured and precise responses, which is likely a result of prior instruction.

Disease Spread and Population Dynamics (Session 3)- Challenge Problem 1: COVID-19 Model – Disease Spread. This section explores the types of ideas generated by the students of both DI and PF groups while working with the models. Based on the responses, this theme is divided into three sub-themes as follows:

1. Direct & Indirect Transmission Pathways: explores how several factors affect disease spread.
2. Public Health Policies & Behavioural Changes
3. Long-Term Immunity & Reinfection Cycles

Direct and Indirect Transmission Pathways

While discussing the COVID-19 model, both groups identified direct contact, respiratory droplets, and surface contamination as primary pathways for COVID-19 transmission. PF students explored different factors affecting disease spread, but some responses show lack of clarity or irrelevant concepts. According to a study, "COVID-19 spreads in different ways. It might spread by contact directly with patient or indirectly by touching things that have the virus." Some of the students highlighted behavioural changes due to the pandemic, demonstrating a knowledge of social and cultural effects, "Shaking hands is common, but when people know that it spreads disease, handshaking decreases in these populations."

In contrast to the PF group, DI students, after receiving direct instructions, gave more structured responses and included scientific terminology: "COVID-19 spreads through respiratory droplets when infected person coughs, sneezes, or talks. People can inhale these droplets or touch surfaces." They also noted biological factors influencing transmission: "The COVID-19 virus has high mutation rate. People who were immune, they become susceptible again because of the new version."

In conclusion, it was noted that PF students explored diverse concepts such as the impact of human behaviour on disease transmission, but their explanations lacked scientific accuracy and frequently omitted essential information through ambiguous reasoning. Their creativity was evident, but their ideas lacked grounding in scientific evidence. DI students, on the other hand, provided systematic and intelligible explanations while correctly applying scientific vocabulary. Their understanding of the subject matter was strong, yet they failed to propose numerous unique ideas and preferred to remain within the bounds of established knowledge rather than experimenting with innovative concepts.

Public Health Policies and Behavioural Changes

The findings indicate that students proposed different methods to control disease spread. It was noted that DI students focused more on public health strategies while PF students discussed individual behaviour more. PF students highlighted the role of personal choices and public awareness while discussing the spread of disease. According to a student, "People's behaviour can reduce spread of COVID-19 spread by wearing masks, keeping physical distance, and washing hands." They also recognised political and government initiatives as influencing factors: "Lockdowns and travel restrictions affect how reduce the spread of the virus."

DI students, in contrast, focused on structural issues, such as healthcare capacity and vaccination coverage: "The availability of vaccines and hospital capacity affects how well can control the disease." Another student highlighted, "If public health policies are not followed, the disease continues to spread, especially in crowded areas with poor ventilation."

Overall, PF students were creative and explored a variety of concepts, particularly those concerning behaviour and policy, although their explanations were not always clear or well-structured. DI students, on the other hand, provided clearer and more organised responses, focused on healthcare systems and large-scale solutions, but they offered fewer innovative ideas. This demonstrates that PF students were better at brainstorming, whereas DI students understood systematic reasoning and scientific details.

Long-Term Immunity and Reinfection Cycles

Some students in both groups agreed that reinfection can occur and that immunity following a COVID-19 infection is not usually lasting. This demonstrated a limited comprehension of the fundamental mechanism of susceptibility and immunity over time in the SEIRS model. A PF group student, for instance, observed: "In COVID-19, the immunity period is for a specific time, which means someone can get the infection again, which means further spread and time to end the disease."

Similarly, a DI student reflected a related understanding but with less clarity: "In the SEIRS model, no one is 100% immune, but you will always be susceptible." These excerpts show that although both groups discussed the idea of recurring susceptibility, DI students used model-specific language to convey it, whereas PF students preferred to articulate it with more real-world context. All of these answers show how students are beginning to understand how reinfection cycles affect a disease's ability to persist in a population.

Challenge Problem 2: (Wolf-Sheep Model – Population Dynamics). This section explores the diversity of ideas generated by the students of both DI and PF groups while working with the models:

- Predator-Prey Interactions
- Ecological Equilibrium & Resource Availability
- Impact of External Factors

Predator-Prey Interactions

While working with the wolf-sheep model, several students were able to recognise the cyclical relationships between predators and prey. The students from DI group noted the basic mechanism of interaction: "Wolves run around eating sheep to reproduce and this consume energy. When a wolf runs out of energy, it dies." Similarly, in the PF group, students were able to identify the predator-prey interaction clearly, with references to the hierarchy within the model: "Wolves prey on sheep and sheep prey on grass."
Another student stated, "There is opposite relationship between grass and sheep and also between sheep and wolves. If one part increases, the other part will decrease."

The references above show students' understanding of how population sizes in the Wolf-Sheep model are interconnected, it shows that the increase in one population (e.g., sheep) leads to a subsequent rise in predator numbers (wolves), which then reduces the number of population (prey) through predation. This decline in prey also affects the predators' survival and reproduction due to limited food resources, which then leads to a decrease in their numbers. The cyclical pattern is then maintained as the procedure enables prey populations to recover. Such explanations indicate that students were starting to grasp the emergent feedback loops that define predator-prey systems.

Ecological Equilibrium and Resource Availability

While discussing the Wolf-Sheep model, both groups identified the idea of balance in the system. DI participants mentioned equilibrium in terms of ecosystem resilience and interdependent resource availability: "There is a kind of balance, as goats eat grass, the number increases, so the amount of grass decreases, but the number of wolves increases because their food has increased."

In the experimental group as well, PF group responses revealed that they were starting to understand how the availability of grass and the reproductive habits of wolves and sheep support ecological balance: "First, there are a lot of these creatures, and it continues to increase as the environment helps this happen. As the sheep eat all the grass in one place, there is grass growing in another one." Another student shared, "If one of them decreases in number, the other one will decrease because of starvation."

These references show that PF students were thinking about how changes in resources like grass being eaten and then growing back, can affect the number of animals that live in a system. Their answers show that they were trying to figure out how the system stays balanced over time, even though they had not been taught about this idea yet.

Impact of External Factors

Students in both groups occasionally noticed how external or environmental factors influence the dynamics of the wolf-sheep ecosystem. One DI group student, for example, observed how natural influences such as wind could impact grass regrowth, indirectly affecting population cycles: "As the sheep eat all the grass in one area, there is grass growing in another area. When the wind comes, it will help the seeds to be transmitted to other places."

One of the students from the experimental group observed the broader role of animal movement and environmental adaptation, “Both sheep and wolves are creatures that move around searching for food and suitable habitats.” (PF)

These quotes demonstrate that even in the absence of explicit instruction as in the PF scenario, and despite varied levels of depth, students were able to examine how aspects outside the immediate model such as wind, habitat dispersion, and mobility can have a substantial impact on species survival and ecosystem consequences.

Struggle (Theme 2). In this section, challenges and struggles faced by students in the Productive Failure (PF) and Direct Instruction (DI) groups are explored.

Session 1. Based on students’ responses, the struggles are categorised into the following sub-themes:

Conceptual Struggles

This theme discusses the difficulties or challenges students faced in understanding reinfection, immunity cycles, and disease spread during the study. The study’s findings indicate that many PF students struggled to clearly articulate how reinfection and immunity cycles work within the SEIRS model (See Appendix J). According to the findings, several students from the experimental group, misunderstood the progression of susceptibility and immune status or made overly general claims about mutation. One of the students from the PF Group shares:

“People not exposed to the disease before, so their immunity against the disease is low… some diseases infect others before symptoms appear, which makes isolation less effective.” Another said: “It is just a computer experiment, not like in real life.”

These quotes show that PF students were reasoning about the nature of immunity and exposure but lacked a clear understanding of how these ideas worked out in the model.

In comparison to the PF group, DI Group students struggled to describe essential concepts such as the cyclic nature of immunity and the impacts of mutation, even after instruction. One of the students from the DI Group stated, “As more people get infected, the immunity of the population increases, which means less infection. Then these people will become susceptible again, meaning the infection will increase again (It is like waves).”

Another DI student talks about it as, “The mutation rate of the disease can make people who are immune become susceptible again, even if they are immune. So, this makes immune periods useless in predicting when the infection would end.”

While the second quote tries to connect mutation with reinfection, it misunderstands the structured phases of immunity within the SEIRS framework.

Perceptual & Model Interpretation Struggles

In this theme, the students discussed the issues in interpreting the SEIRS and Forest models, including visual illusions and simulation mechanics. When engaging with the forest fire model, PF students often assumed the fire had actual movement or confused it with biological spread. Many focused on external elements like wind or resistance instead of understanding state change.

Some of the references from the PF Group are: “Fire does not move until another factor comes to increase its ability to move, like wind”, “It's clear to us that fire is moving, but it fights to survive” and “The fire here follows the same concept as disease spread. The disease does not move from one person to another, but spreads from the sources.”

These comments show the misinterpretations of the model's dynamics as well as confusion between various sorts of emergent behaviour. DI Group students, on the other hand, struggled to understand the visual illusion and provided overly complicated or abstract answers. One of the students from the DI Group shared: “The simulation could use visual effects like color changes or effects to show the fire spreading, though the pixels themselves are not moving." Another stated, "The pixels are not moving as they are already set, so the fire is not spreading when burning the trees."

Here, DI students appear to understand the illusion but end up overcomplicating their explanations, missing the opportunity to explain emergent behaviour more directly.

Reasoning & Explanation Struggles

This theme focuses on overgeneralisations, inconsistencies, and missing essential causal links in student responses. It emphasises the difficulty students have in making obvious links between model behaviour and real-world issues. The findings revealed that when comparing SEIRS and Forest Fire models, PF students frequently provided surface-level comparisons or ambiguous explanations. A student from PF Group shares her views as, "They represent the same concepts, just demonstrated in different ways."

Similarly, other students talked about it as, "Both models represent the dangers of infectious disease and how it spreads" and "In the SEIRS model, the infection prevalence rate may be affected by several factors… but in the tree model, trees don't recover."

These responses show an effort to compare the models but reveal confusion about the distinct roles of recovery, immunity, and agent behaviour. DI Group students also showed difficulty remembering or distinguishing model features. Students from DI Group shared their views as: "I can't remember how SEIRS and Forest Fire models are related" and "the SEIRS model is more accurate than the tree model, as it looks at the spread of the disease from many aspects rather than mainly one."

However, the last quote fails to highlight the fundamental difference in model structure (fire as an irreversible state change vs. SEIRS as cyclical), demonstrating faulty reasoning and inadequate model-based explanation.

Session 2. This section examines the challenges students faced while working with the following challenge problems:

- The Malaria Transmission ABM, which focuses on vector-based disease spread and healthcare system capacity.
- The Marketing ABM, which explores consumer decision-making, tipping points, and product persistence in the market.

Based on the responses from students of both groups, this theme is divided into three sub themes, which include Conceptual Struggles, Perceptual & Model Interpretation Struggles and Reasoning & Explanation Struggles. Each of these themes is discussed below.

Conceptual Struggles

PF students together with DI students found it challenging to understand the complete process of malaria transmission, especially the cycle involving human-to-mosquito-to-human transmission. A few students proposed the incorrect idea that malaria transmission could occur without the involvement of a vector.

According to a PF group member, "The disease spreads and starts in a specific area where there is a vector that spreads it or spreads without a vector." Despite receiving previous lessons, the DI students demonstrated confusion about how reinfection occurs and how disease spreads over time: "Awareness can affect how people prevent the disease and how it spreads, but if resources are not available, prevention becomes difficult."

PF students who misunderstood hospital capacity's impact on ICU occupancy could not clarify how expanding hospital capacity leads to a decrease in severe cases. According to a PF student, "I have noticed that when the capacity of the hospital or the ICU increases, the number and ratio of patients being hospitalised or in the ICU decreases, which I couldn't understand why."

In conclusion, PF students struggled to understand indirect disease transmission and hospital operations during outbreaks. Connecting different aspects proved difficult for them, such as understanding the environmental transmission of infections and hospital patient and resource management systems. On the other hand, DI students demonstrated a better-organised comprehension which enabled them to describe the mechanisms of disease transmission and hospital functionality more effectively. The DI students occasionally failed to provide detailed explanations about reinfection processes including the roles of immunity changes, persistent pathogens, and human mobility in initiating new disease outbreaks (See Appendix J).

Perceptual & Model Interpretation Struggles

Both student groups struggled to understand how Marketing ABM represented customer decision-making and tipping moments. Several PF students failed to appreciate how product quality influences adoption patterns because they ignored system-level interactions. A PF group student states, "The quality can determine if the product will stay in the market. Many products have disappeared when finding better alternative, even if the cost is high."

DI students demonstrated a better awareness of how tipping points affect product uptake, but they struggled to identify the fundamental indicator that influences consumer purchasing decisions. A student opines, "The best method that controls whether people decide to never buy a product or not is having the short attention span."

Additionally, students had difficulty explaining the relationship between memory, forgetting, and product cycles because PF students saw forgetting negatively rather than as a re-engagement technique. According to a student, "It is helpful for a population to forget the product they purchase, because they may change their mind when they see it again. So there is a cycle pattern with the never-buy parameter."

In conclusion, PF students faced difficulty following consumer behaviour change because they tended to focus on current decisions rather than future trends. Students failed to connect changes in preferences and market impacts to external factors such as economic situations or technological breakthroughs that influence consumer behaviour patterns. Their responses revealed insufficient comprehension of how historical experiences together with evolving societal norms influence behaviour patterns. On the other hand, DI students better understood structured decision-making which enabled them to present clear logical steps in consumer options. The students struggled to pinpoint the main variables which drive behaviour including psychological elements and social or economic trends. Students provided descriptions of how decisions are made but failed to identify the underlying factors that drive consumer behaviour. DI students demonstrated a better awareness of how tipping points effect product uptake, but they struggled to identify the fundamental indicator that influences consumer purchasing decisions.

Session 3. This section examines the challenges students faced while working with:

- **The COVID-19 Transmission ABM**, which focuses on exploring the impact of factors like social behaviour, public policies, and healthcare infrastructure on the spread of the virus.
- **The Wolf-Sheep Predation Model**: It focuses on how predator-prey relationships impact population dynamics as well as ecological balance.

Based on the students' responses from both groups, the challenges or struggles are categorised into three sub-themes which include conceptual struggles, perceptual & model Interpretation struggles, and reasoning & explanation Struggles.

Conceptual Struggles

The findings indicate that students in both groups faced difficulties in understanding concepts like reinfection, immunity loss, and public health responses in the COVID-19 model

(See Appendix J). The results show that some of the PF students failed to understand how immunity works, which led to confusion about why the disease continues to spread. According to a PF student, "COVID-19 has an incubation period before symptoms appear, which may contribute to spreading the disease further. But I don’t get how some people are asymptomatic even after the incubation period and still spread the disease." Responses also show that some of the students also struggled to understand the impact of policy measures, questioning how government interventions affect infection rates over time: "Government policies like lockdowns and restrictions slow the spread, but why do cases increase again after the lockdown is lifted?"

Similarly, students from DI group, despite getting formal instructions, were also found to be struggling with conceptual difficulties particularly in understanding long-term immunity and virus mutations. A DI student says, "The high mutation rate of the virus makes people who were immune susceptible again. But does that mean vaccines won’t work in the long run?"

In the Wolf-sheep model as well, students from both groups were found to be struggling with the idea of the tipping point. It was noticed that some of the students from both groups struggled to explain when and why population levels shift dramatically in the Wolf-Sheep model. The analysis of the responses shows that the tipping points, such as sudden collapses or stabilisations, were sometimes misunderstood or oversimplified by the students.

It was noticed that some PF students found it difficult to predict when a tipping point would occur. A PF student says, “The number of sheep and wolves always stays the same, no matter what.” Similarly, a student from the DI Group states, “If too many sheep are eaten, they all disappear, but the wolves don’t die.” These references demonstrate a lack of understanding of how interdependent species cycle’s function and how resource shortages or over-predation can generate nonlinear population fluctuations. Students frequently fail to recognise important thresholds or feedback processes that trigger these shifts i.e. true tipping points in the system.

Perceptual & Model Interpretation Struggles

The results indicate that students from both groups showed similar challenges when attempting to understand how the ABMs modelled real-world processes. PF students misunderstood the function of asymptomatic carriers in the COVID-19 model because they thought every infection should lead to noticeable symptoms. According to a PF student, "Some people can spread the disease without knowing they have it, but do they feel sick at some point?"

Similarly, DI students had trouble understanding the link between human behaviour and virus transmission rates and questioned the reasons behind small movement pattern changes impacting the whole population as according to a student, "If only a small number of people ignore health guidelines, why does the infection rate rise so quickly?". Furthermore, the analysis also indicates that the students were also found to be struggling with interpreting the population dynamics in the Wolf-Sheep model, especially how population balance is maintained through interdependent predator-prey relationships. One of the PF group students shared: "Wolves hunt sheep until there are none left, then they die too." Similarly, a student from the DI group explained, "If sheep keep increasing, wolves stop hunting and eat grass instead."

These interpretations from the students reveal that students misunderstood the simulation mechanics. It was noticed that they either assume that wolves keep eating without limits or behave in ways that do not match how predators and prey normally interact in the model. Such misconceptions highlight perceptual and conceptual struggles with ecological equilibrium and system logic.

Reasoning & Explanation Struggles

The results also show students having difficulty in reasoning as some students were found to be able to recognise key factors in disease spread and marketing adoption but struggled to structure their explanations logically. For example, in the COVID-19 ABM, a PF student noticed how behavioural patterns influence transmission but failed to clearly explain the connection:

PF Group: *"COVID-19 changed how people interact and become more cautious. But I am still thinking why the infection rate parameter goes down only after long time."*

Similarly, DI students, despite getting proper instructions, were mostly presenting general statements without deeper reasoning:

DI Group: *"If people follow public health measures, the disease should go away. But I guess it doesn't always work like that."*

In the Wolf-sheep model as well, several students failed to recognise or clearly explain critical tipping points, such as population crashes or sudden stability in predator-prey interactions. These moments represent complex feedback loops often missed in student responses. A student from the PF group shares, "The number of sheep and wolves always stays the same, no matter what."

Similarly, a DI student quotes: "If too many sheep are eaten, they all disappear, but the wolves don't die." These quotes reflect students' lack of understanding of nonlinear system behavior. These quotes show that students were unable to fully understand how predator-prey systems can change a lot from small changes in conditions or external influences

Relevance of Ideas (Theme 3). This theme evaluates the relevance of ideas based on the students' responses from both groups PF and DI.

Session 1. This section examines the relevance of students' responses in explaining the spread of disease in the SEIRS model and fire in the Forest Fire model during Session 1. The responses from the PF (Experimental) and DI (Control) groups are compared to see how well their explanations aligned with the actual model mechanisms.

Based on the responses, this theme is divided into sub-themes, which include:

1. **Understanding of Core Model Mechanics**: How well students explained the key processes of each model.
2. **Appropriateness of Analogies & Real-World Connections**: How effectively students related models to real-world phenomena.

In this session, students were asked to explain why disease continues spreading over time using the SEIRS epidemic model.

Understanding of Core Model Mechanics

The findings of the study indicate that both PF and DI students demonstrated the ability to explain core principles of immunity loss and reinfection within the SEIRS model. Students understood the progression from vulnerability through exposure and infection to recovery while also acknowledging that people can experience reinfection after their immunity diminishes. A group of students demonstrated accurate understanding of these processes, but some students had difficulty with the concept of reinfection and missed essential information about how immunity develops over time. Here are some instances of student responses: "As seen in the model, the spread of the disease depends on many factors, such as the number of people contacted per day and the infection rate, to what extent this virus can be transmitted to others." (PF Group) and "The high mutation rate of the disease can make people who are immune susceptible again, even if they are still immune to the older variant." (DI Group).

PF students discussed human interactions and transmission rates while DI students examined biological mutation and reinfection cycles in their relevant responses. Furthermore, some students demonstrated their understanding of complex systems by noting that fire spreads through state transitions in each tree rather than physical movement. Students' replies demonstrate understanding of the concept that fire propagation occurs via a succession of local tree state changes rather than physical movement. According to a PF group student, "The fire is not moving from place to another but spreading in different directions… it can spread to other trees, which changes and show us like moving fire." On the other hand, a DI group student states, "The simulation could use visual effects like color changes or effects to show the fire spreading, though the pixels themselves are not moving."

While both groups provided relevant explanations, PF students focused on system interactions, whereas DI students interpreted the fire spread as a visual illusion created by the simulation. Some students attempted to explain fire propagation using analogies from other fields, specifically epidemiology and visual effects in media. While these parallels exhibited creative thinking and an attempt to connect the model to recognisable concepts, they did not fully explain the physics of fire movement in the simulation.

Appropriateness of Analogies & Real-World Connections

A number of students attempted to interpret the SEIRS model's application to practical public health problems including insufficient disease prevention efforts and policy failures. Students demonstrated their ability to translate theoretical knowledge into practical applications through these connections which helps to understand how epidemiological models guide public health policy decisions.

The answers provided insufficient details and did not clearly show how real-world public health problems relate to SEIRS model mechanics. Students analysed public health failures through broad descriptions instead of linking to precise simulation parameters like immunity duration and infection probability. For example, a PF student mentioned the role of poor prevention and lack of pathogen knowledge in disease spread: "Bad prevention affects the infection and spreading of disease… but don't know the type of pathogen." Similarly, a DI student made a general statement about public health measures but did not explain how this connects to the SEIRS model, "The failure of public health measures allows the disease to continue spreading." Both responses highlight important public health concerns, but neither fully explains how these factors interact with the SEIRS model parameters, such as reinfection rates or immunity loss.

Session 2. This section examines the relevance of student responses when engaging with:

- The Malaria Transmission ABM, which focuses on disease prevention, hospital capacity, and vector management.
- The Marketing ABM, which aims to identify tipping points in customer behaviour and product uptake.

Based on the responses, this theme is divided into two subthemes which are:

1. Understanding of Core Model Mechanics: Discusses how well students understood malaria transmission and product acceptance.
2. Appropriateness of Analogies & Real-World Connections: Discusses students' ability to effectively apply models to real-world scenarios.

Understanding of Core Model Mechanics

The findings of the study show that although both groups understood malaria transmission through vectors, they varied in their ability to explain what influences transmission. PF students explained important transmission pathways, yet their explanations remained general without mentioning specific system interactions. A student says, "Malaria spreads through the vector, which are mosquitoes, by moving between people. Many factors affect the spread, including a long incubation period and low immunity." They pointed out that epidemics overwhelm hospitals but experienced difficulty in directly linking hospital capacity with disease management: "A high fraction of people requiring hospitalisation or ICU indicates that hospitals and ICUs are full, therefore more deaths."

Through direct instruction, DI students delivered more structured explanations and demonstrated better comprehension of tipping points in healthcare capacity. A student says, "If a large number of mosquitoes become infected with malaria, the spread of the disease may reach a tipping point where the number of cases increases rapidly." Another student states, "Tipping points can also be seen in hospital capacity. If ICU admissions exceed available beds, deaths increase significantly."

PF students demonstrated comprehensive knowledge of the subject and proposed diverse ideas yet faced difficulties when required to apply structured cause-and-effect reasoning. Respondents did not consistently demonstrate clear linkages between different factors which made it hard to follow how certain circumstances produced specific results such as epidemic development or healthcare intervention results. In contrast to the other group, DI students provided more specific explanations that linked hospital capacity to important epidemic tipping moments. They displayed systematic comprehension by explaining how patient overflow, resource constraints, and delayed responses bring healthcare systems to a crucial breaking point.

Appropriateness of Analogies & Real-World Connections

While both groups utilised real-world comparisons to describe concepts, PF students employed fewer accurate analogies compared to DI students who demonstrated stronger conceptual links. PF students equated malaria transmission to daily social exchanges, though their analogies were more applicable to direct-contact diseases rather than vector-borne transmission:

Some PF students equated malaria spread to everyday social interactions: "People’s normal behaviour, like visiting others, shaking hands, and hugging can lead to spread the disease more." DI students demonstrated a more accurate analogy of malaria spread by associating it with poverty-related structural issues that directly influence disease transmission. According to a DI student, "The poor people can’t afford to move from malaria areas. This means they are highly affected due to economic issues, and this will leading to outbreaks."

In the Marketing ABM, PF students focused on practical marketing strategies, while DI students applied the concept of tipping points more accurately, "If we reach of a large group through easy marketing channels (e.g., Instagram), this can increase product adoption." Another DI student says, "Tipping points occur in marketing when a product reaches a high level of acceptance and increase adoption."

In conclusion, both PF and DI students demonstrated their ability to link learning material to real-world situations but used different approaches to apply the models to the real-world scenarios. It was observed that PF students explored multiple perspectives but struggled with systematic reasoning and understanding complex systems. In contrast to PF, DI students who received prior instruction demonstrated clearer explanations but occasionally overlooked fine details in behavioural interactions. These findings align with Productive Failure theory, where PF students produce numerous unorganised ideas before receiving instructional materials whereas DI students depend on structured learning but often fail to discover wider conceptual linkages.

Session 3. This section evaluates the degree to which students' responses matched the concepts they studied during the session. These models include:

1. The COVID-19 Transmission ABM: It demonstrates the impact of social behaviours, public policies, and healthcare interventions on virus transmission.

2. The Wolf-Sheep Predation Model: It focuses on how predator-prey relationships impact population dynamics as well as ecological balance.

To evaluate how well students applied their knowledge and reasoning, their responses were categorised into two sub-themes which include: Understanding of Core Model Mechanics and Appropriateness of Analogies & Real-World Connections.

Understanding of Core Model Mechanics

The results indicate that both groups demonstrated an understanding of how COVID-19 spreads. It was noticed that the PF students focused more on behavioural patterns in COVID-19 spread while DI students provided structured explanations focused on biological mechanisms. It was noticed that PF students were able to describe how human actions influenced virus transmission but often lacked precise scientific reasoning to support their views. A PF group student states, "Shaking hands is a tradition, but when people realized this could spread diseases, handshaking decreased. We saw this happen during COVID-19—reducing contact reduced the spread over time." Some PF students attempted to relate disease spread to concepts from other models, such as fire propagation: "Like in the fire model, the more people an infected person comes into contact with, the more the disease spreads."

While this analogy captures the idea of rapid transmission, it fails to account for differences in disease contagion and spread dynamics, such as reinfection and immunity. DI students, in contrast, provided more structured, factually accurate responses: "COVID-19 spreads through respiratory droplets. If public health policies are not followed, the disease continues to spread, especially in crowded areas with poor ventilation." and “The SEIRS model represents waves of infection and reinfection because immunity is not permanent. This is different from the fire model, where the spread follows a one-directional pattern."

Appropriateness of Analogies & Real-World Connections

Both groups tried to make connections to the real world, but DI students showed a closer conceptual alignment while PF students often applied analogies loosely. Some PF students compared COVID-19 spread to fire spread, though this analogy missed key important differences between disease transmission and physical combustion. A student says, "COVID-19 spreads like fire the more people it touches, the more it spreads."

DI students understood the distinct mechanisms behind fire propagation and disease transmission. For instance, "In the SEIRS model, people can recover and become susceptible again, leading to cycles of reinfection. In the fire model, burned areas do not catch fire again."

In the Wolf-Sheep Model as well, PF students made general observations about predator-prey interactions, while DI students provided more technical explanations. The PF student states, "Sheep eat grass, wolves eat sheep, and the cycle continues." On the other hand, the DI student says, "The wolf-sheep model represents dynamic equilibrium—population numbers fluctuate based on resource availability and predation rates."

The research shows that PF and DI students demonstrated success in connecting models to real-world situations through distinct methods. PF students investigated multiple perspectives which allowed them to develop a range of ideas concerning the spread of disease as well as consumer behaviour and ecological systems. The students produced insightful explanations, but their reasoning structure was weak and analogical precision was missing so their explanations often diverged from the models' fundamental mechanics. Students who had prior instruction delivered structured answers with factual accuracy that demonstrated their knowledge in epidemiological patterns along with marketing strategies and predator-prey relationships. Their structured knowledge base constrained their ability to develop broader conceptual insights and explore alternative explanations outside established frameworks.

The findings support the productive failure theory, which argues that PF students meet initial difficulties but attain deep comprehension through exploratory learning while DI learners get instructed information but demonstrate weak engagement with profound conceptual comprehension. Analysis of the learning methods demonstrates their benefits as well as restrictions while highlighting the necessity to unite self-regulated learning with the structured support to understand the intricate system.

Analysis of Challenge Problem 3 Data

In this section, analyses of the two tasks associated with Challenge Problem 3 are presented. First, a statistical analysis was performed on the data from the application task. This was done to determine whether the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemics concepts affected the learning process across multiple sessions. Next, a qualitative analysis based on a specific coding scheme was performed on the data from Challenge Problem 3's compare-contrast task.

Analysis of the Application Task. For reasons discussed in Chapter Four, the data from the application task were analysed Challenge Problem 3. This problem was particularly important for the experimental (PF) group, since it was placed after the exploration phase and the guided learning videos. Given the violation of the normality assumption documented in Chapter Three, in order to determine whether there were significant differences in the two groups' scores on the application task of Challenge Problem 3, Mann-Whitney U tests were run on each session's data between the PF and DI conditions. The results of the Mann-Whitney U test of the application task scores for each session are presented in Table 25.

Table 25

The results of Mann-Whitney U test in each session (PF vs. DI)

	Session 1	**Session 2**	**Session 3**
Mann-Whitney U	120.000	125.500	103.500
Z	-1.057	-0.900	-1.687
R	-0.179	-0.152	-0.285
p-value	0.291	0.368	0.092

No significant difference was found between PF and DI groups in the mean ranks of session 1 (*U* = 120.00, *p* =.291, *r* = .179) or session 2 (*U* = 125.50, *p* =.368, *r* = .152). Here it is important to note, however, that the mean ranks of the experimental group in both sessions were greater than the control group.

The difference in the mean ranks increased in session 3 between PF (mean rank = 20.33) and DI (mean rank = 14.90) groups, a difference that was nonetheless not significant (*U* = 103.50, *p* = 0.092, *r* = .285). As *r* follows the same threshold values as Cohen's *d* (1998), the effect size here is approximately 0.30, indicating it is approaching a medium effect in the context of the Mann-Whitney U test. A small effect size is classified as *r* =0.10, while a large effect size is *r* =0.50. For the Challenge Problem 3 score across all 3 sessions (i.e., the average score of the 3 sessions), an independent samples *t*-test was run, as the normal distribution assumption, examined via a Shapiro-Wilk test, was not violated here. The results are presented in Table 26

Table 26

The Results of Independent Samples t-test for all sessions (PF vs. DI)

Outcome	**t**	**p-value**	**Cohen's d**
Problem 3 Rating Score (All Sessions' average)	1.658	0.107	0.566

No statistically significant difference was found in the mean Challenge Problem 3 scores between the PF and DI groups (*t (DF)* = 1.658, *p* = 0.107, *d* = 0.566). However, the mean rating score of the experimental PF group (M = 2.15) is higher than the control DI group (M = 1.78). Moreover, the effects size *d* < 0.5 can be considered a 'medium effect' for a *t*-test. This suggests that the observed effect is meaningful and may be worth considering.

Analysis of the Compare-Contrast Task.

In the last phase of each session, students were presented with a compare-contrast task requiring them to note similarities or differences between the two models they had used earlier in that session. The compare-contrast tasks were based on the Analogical Comparison (AC) and Productive Failure (PF) learning approaches. These methods assist students in conducting independent investigations of concepts prior to receiving structured explanations. A coding scheme for the compare-contrast task was adapted from two established frameworks (Gentner et al., 2003; Jacobson et al., 2020). Specifically, the scheme used in this study focused on two key dimensions: surface-level features and structural-level features.

Session 1. This section presents the results of Session 1, in which students were asked to compare the SEIRS epidemic model and the Forest Fire model. The major goal of this session was to investigate how students perceived similarities and differences between the two models, with a focus on spread dynamics, complexity, and system behaviours (See Appendix J).

The findings are categorised into two primary themes, each of which includes numerous sub-themes that pertain to particular aspects of the comparison:

Structural-Level Features (six sub-themes) - These are the fundamental mechanisms that govern each model, including causal relationships, feedback loops, and emergent behaviours.

Surface-Level features (one sub-theme): This theme has only one sub-theme, which focuses on the visible and visual aspects of models, such as how data is displayed and interpreted.

This section provides insights into how students perceived the two models differently and how well they understood the complexities of fire spread and disease transmission by analysing the responses of students in both control and experimental groups

Structural-Level Features

Emergent Properties

Both models are characterised by the phenomenon of emergence, which is the consequence of small-scale interactions that lead to large-scale behaviours. A student in the control group expressed a general acknowledgement of this, stating, "They both look at emergent properties of something that spreads."

On the other hand, the experimental group demonstrated a more profound understanding of this concept, clearly recognising the influence of system parameters on emergent behaviours. One student noted, "Both models are examples of complex systems. They demonstrate emergences and contain important parameters that control the microlevel of the system." Another student provided more detail about how emergence is self-regulatory and how distinct restrictions affect how the system acts: "The way that fire spreads among trees is similar to the way that infection spreads from person to another. Both models contain parameters that prevent emergence to happen like tree heat resistance or infection period."

In conclusion, although both groups agreed that emergence was a key trait, the experimental group had a clearer picture of how it is related to specific factors like immunity, latency, and resistance.

Factors Influencing Spread

In this theme, the students identified factors that can influence the spread dynamics. Students in both groups identified significant aspects influencing spread dynamics, but the level of detail in their replies was observed to be varied. A student from the control group noted, "All of them depend on the main factor that may affect the spread. Fire depends on the distance, and the SEIRS model depends on the characteristics of the disease."

In contrast, a student from the experimental group provided a more complex and layered analysis, considering additional variables beyond spatial factors.

In both models, the SEIRS model and the forest fire model, the number of objects increases when the density of and closeness to each other increase. However, in the SEIRS model, the infection prevalence rate may be affected by factors like the infectious organism.

The findings of the study indicate that the students from the experimental group included biological aspects like immunity, infection strength, and latency in their explanation while comparing these models. However, the control group, in contrast, mostly focused on proximity as the distinguishing factor which shows a higher-level understanding of disease transmission.

Model Complexity and Outcomes

In terms of model complexity and outcomes, both groups were able to recognise the differences in both models. While discussing the comparison, both groups were of the view that SEIRS model is more complex than the Forest Fire model, but the findings indicate that the experimental group had a better knowledge of working with parameters. They shared how particular parameters could be changed to impact the model's behaviour. One of the control group students discussed the Forest Fire model as a straightforward, one-way process, stating, "Forest fire model is a one-way process with one outcome, but the SEIRS model has multiple outcomes."

The quote above shows that students understand the difference between those models. They understand the concept that fire spreads in a linear, irreversible fashion, but disease transmission occurs in more complex cycles, allowing for reinfection and recovery. However, the control group responses lacked specificity as to how these changes manifested in the models.

In contrast, students in the experimental group engaged in a more nuanced debate of model complexity. One student explained, "In the SEIRS Model, there is more variables to test, like the incubation period."

This shows that students in the experimental group were more aware of SEIRS's adjustable components, such as how changing the incubation period, infection likelihood, or immunity length might drastically alter the model's results. Another student offered more comparative information, explaining, "The SEIRS model assumes that no one dies, so there are always new people to infect if the conditions are ok, whereas in the forest model, fire spread is permanent."

This direct comparison reveals a better conceptual understanding of how each model manages resource renewal: "although in the Forest Fire model if a tree burns, it cannot recover, in SEIRS new susceptible individuals constantly enter the system". Although both groups agreed that SEIRS is the more complicated model, the control group offered only surface-level observations while the experimental group showed a deeper knowledge by realising the adjustable parameters and long-term dynamics that separate SEIRS from the fixed, irreversible character of the Forest Fire model.

Immunity Differences

The results show that both groups were able to recognise that immunity plays a key role in the SEIRS model but is absent in the Forest Fire model, making it a key distinction between the two. This distinction was acknowledged by the students from the controlled group, but they too were unable to analyse it in depth. One student simply noted, "Fire spread is irreversible, while SEIRS includes immunity and recovery."

This shows that students had a fundamental understanding that the SEIRS model has an evolutionary system in which individuals can recover and potentially build immunity, which the other model cannot do as it irreversibly burns fuel, limiting regeneration within the model's timescale. In contrast to the control group, students in the experimental group show a deeper understanding of how immunity functions over time. One student compared the two models by stating: "The more the resistance of trees or the more immunity of people, the less the spread of the disease or fire. In both models, when the density of people or trees increases, the fir or disease spreads more."

This quote shows that both models have mechanisms that can slow or accelerate spread, whether through natural resistance such as tree density, fire resistance in the forest fire model or acquired immunity such as vaccine, infection recovery in the disease model. However, students in the experimental group displayed a more in-depth understanding of how immunity works overtime. One student compared the two models and stated, “The SEIRS model allows reinfection and has a latency period where no transmission occurs.”

Another student elaborated on how immunity affects long-term outcomes stating, "The SEIRS model assumes there are always new people to infect as no one dies, whereas in the forest model, once tree burns, it does not recover." This shows the major difference between the two models. As shown in the quote above in SEIRS model, the population is always changing as people lose their immunity and new individuals come into the population who are more likely to get sick. But in the Forest Fire model, things move more slowly, and once damage is done, it's done for good. Overall, the experimental group showed a greater understanding of how immunity limits disease progression over time, whereas the control group recognised the concept but did not investigate its broader implications.

Transmission and Repeatability

Similarly, while discussing transmission and repeatability, students from both groups were able to identify that transmission dynamics differed significantly between the two models. However, it was observed that the experimental group provided a more detailed analysis of how these differences influenced system behaviour. A significant distinction between the SEIRS and Forest Fire models is the capacity of a system to repeat or sustain transmission over time. In contrast to fire, which spread in a single irreversible manner, disease in the SEIRS model follows a cyclical pattern in which individuals may re-infect themselves after being immune for a period. This distinction has significant implications for the long-term behaviour of each system, particularly in the areas of system stability, intervention, and containment.

Students in the control group identified this fundamental difference but provided a surface-level description of repeatability. One student stated, "Once a tree is burned, it cannot be burned again, so no repeating cycle like SEIRS." The study findings indicate that students from the control group did not go into detail on how immunity loss or external interventions affect repeatability, instead, they focused more on recognising the difference. However, the experimental group investigated in greater depth how illness reinfection, latency, and immune loss affect system behaviour. A student explained, "The SEIRS model allows reinfection and has a latency period where no transmission occurs at this period." Another student provided further insight into the role of external interventions, stating, "In the SEIRS model, we can control the spread by adjusting factors like immunity and infection probability, whereas fire spread depends on conditions like wind."

This response shows that the students from the experimental group had a deeper understanding of epidemiological concepts by acknowledging that disease transmission is regulated by factors such as immunity and incubation periods. Unlike fire, where once fuel is exhausted, the spread cannot resume, the SEIRS model features multiple pathways that allow the disease to persist within a population. This means that students in the experimental group knew more about how changing factors, like vaccinations and social distance, can change the way diseases spread. On the other hand, students in the control group thought that fire spread was a physical process that could not be changed by outside forces.

Surface-Level Features

The students were also asked to identify surface-level features between the SIERS and the Forest fire models. It was observed during the analysis of the responses of the students that both groups were able to identify basic visual and structural similarities without necessarily identifying deeper causal mechanisms. These comparisons were mostly based on observable characteristics, such as how each model represents spread and how users can manipulate variables. Both groups recognised that the models visually depict the spread of an entity, such as fire or infection. A control group student noted, "The spread of fire in the forest is like the spread of disease among people."

This statement shows that students were able to make an analogy in which students compared the visual patterns of infection spreading through people to fire spreading through trees. This observation, however, ignores the several underlying mechanisms that control each step, such as fire dependence on ambient variables or reinfection in SEIRS.
Similarly, students found basic changeable parameters in both models. A control group student said, "Both have simulation videos, and in both, you can control the number of contacted people (or tree density) and infection probability (tree heat resistance)."

The findings also revealed that although students were able to recognise similar interactive features in both models, such as adjusting variables that influence spread, it was also observed that the students from the control group did not explore how these variables function differently. In contrast, students in the experimental group provided a comparatively more detailed discussion of these features. According to a student, "Both give a visual representation of the spread. In both models, we can adjust some variables."

This was just a basic observation, but some students in the experimental group went into more detail about how data is presented and studied in different ways. One of the students said, "In the SEIRS model, we can view the results in the form of a chart. In the forest model, we can find the ratio."

Overall, both groups found significant similarities in how the models depict spread and how variable changes. The experimental group recognised additional differences in how data is displayed and analysed, while the control group mostly looked at direct visual comparisons. Even so, neither group really got into the deeper structural differences between the models. This supports the idea that observations at the surface level are not enough to understand how complicated a system is.

Session 2. During this session, students were required to apply the notion of tipping points to two different models. These models include the Malaria Model and the Marketing Model. A tipping point is a key threshold at which slight changes cause major adjustments in system behaviour. In this session, students explored how tipping points influence disease transmission in malaria model and consumer behaviour in the marketing model.

Structural-Level Features

Based on the responses of the students, this theme is divided into four sub-themes, which are presented below:

1. Impact of External Conditions: In this theme, students explore the impact of policy or environmental factors on changing the tipping points.
2. Role of Preventive Measures: This theme focuses on measures to delay or avoid tipping points.
3. Self-Sustaining Spread: This theme talks about how breaking points cause a system to keep changing.
4. Thresholds and Critical Mass: This theme explores the conditions or situations which cause a system to change dramatically.

Impact of External Conditions

External conditions, such as policy changes, economic shifts, or environmental factors can influence tipping points. These external factors can prevent the system from reaching its tipping point by either accelerating, delaying, or preventing it. The findings of the study show that the students in the control group were able to recognise that one of these factors i.e. policy changes could reduce the likelihood of tipping points, but they viewed them as fixed thresholds. A control group student noted that sudden policy changes could prevent tipping points, stating, "They could refer to sudden policy change, where the spread is highly decreased."

By contrast, the experimental group provided a more specific explanation, linking external conditions to consumer behaviour in marketing: "A tipping point occurs when the product quality becomes good enough (threshold), leading to increase customer satisfaction and repeat purchases." The students from the experimental group, on the other hand, showed in their responses that tipping points change over time because of how systems interact with each other. They recognise that factors such as immunity, feedback loops, and external influence can cause these points to change, which highlight a deeper understanding of how systems fluctuate between stability and change. The experimental group expanded on this concept, linking external conditions to tipping points in both malaria and marketing. A student states, "The increase or decrease in these parameters will have a dramatic effect on the number of the deaths, infected people, and the immune ones." In another response, it is said, "A tipping point occurs when the product quality becomes good enough (threshold), leading to increase customer satisfaction and repeat purchases."

In summary, students demonstrated an understanding that the dynamics of tipping points can be influenced by external interventions, such as the improvement of healthcare services in malaria or the enhancement of product quality in marketing. This underscores the importance of a more adaptable understanding of the way in which different systems react to external influences.

Role of Preventive Measures

The findings of the study indicate that preventive measures play a crucial role in delaying or avoiding tipping points by changing system parameters before they reach a critical threshold. This is particularly critical in disease models, as early intervention can prevent widespread outbreaks.

One of the students in control the group acknowledged this, stating: "Following the policies of prevention of malaria will help to reduce the risk of malaria."
While this response shows that the students from the control group highlight a basic awareness regarding the factors that can prevent outbreaks, they do not go in-depth to explain how interventions influence the tipping point. The experimental group provided a more structured explanation, linking specific interventions to tipping points: "Tipping points may occur when the incubation period gets too short or too long. This could impact on how to control disease as timing may affect interventions."

This response shows that the student had a deeper comprehension of the relationship between time and system parameters, as he accurately explained that the intervention delays might cause a system to reach its tipping point. Another student connected this to marketing, stating, "The evaluation time (when customers evaluate the product) may reach a tipping point, showing that customers have positive thoughts and are more likely to buy."

This shows that there are preventive measures for both marketing and malaria, such as early advertising to attract consumer interest or early treatment to lower infection rates.
In conclusion, both groups understand the basic concept of tipping points can cause significant changes in how a system works. However, the control group had a basic understanding of these concepts as they were unable to give detailed explanations, while the experimental group showed a better grasp of feedback loops, external factors, and system changes. The experimental group also found stronger links between disease models and

marketing models which shows how tipping points work in various fields.

Self-Sustaining Spread

According to the results, one of the distinguishing features of tipping points is their ability to generate self-sustaining change within a system. In the malaria model, this occurs when infections reach a level where the disease spreads uncontrollably, whereas in the marketing model, it happens when a product achieves extensive popularity, which diminishes the necessity for advertising. In a response, students from the control group shares the general idea they have regarding tipping points under the malaria model, "In the context of the malaria model, a Tipping Point is the point when the spread of the continues spreading on its own within a population."

While this response correctly describes what a tipping point is and how it works, but the findings show that students does not go into more detail about what makes these models self-sustainable. The students in the control group mostly talked about defining the concept instead of discussing the factors that contributed to it. The experimental group, on the other hand, showed a better grasp of feedback loops and how systems work. One of the students says, "The tipping points reach peak at the beginning of the outbreak. As more people get infected, the government will start to control the disease. After some time, they become stable and achieve balance between susceptible, exposed, and infected people."

This response from experimental group students shows how tipping points evolve over time due to system interactions. While the control group perceived tipping points as fixed thresholds, the experimental group explained that they can shift in response to external influences, feedback loops, and immunity.

Thresholds and Critical Mass

A tipping point occurs when a system reaches a threshold where small changes can cause dramatic shifts. In the malaria model, this could be hospitals reaching full capacity, which means it can lead to more deaths. In marketing, it could mean that a product gets bought by so many people that its adoption quickly spreads. One of the students from the control group highlighted this idea in the context of hospital capacity, stating:

> Tipping point means the point that makes a difference in the situation, so in this case, the tipping point will be an increase in the number of infected people so we cannot help them because they are more than the number of hospitals or the beds in the hospital that can treat them.

Although, this response correctly identifies healthcare capacity as a tipping point in the Malaria model that if the hospital capacity reaches its limit it can contribute to higher mortality rates. However, it does not explain how different parameters (e.g., infection rate, immunity, intervention timing) can influence the threshold. In contrast to the control group, the experiment group gave a comprehensive explanation by discussing multiple factors that can influence tipping points: "Tipping points may happen in the context of hospital and ICU capacity. If the number of malaria cases exceeds the capacity of healthcare, a tipping point is reached, which affect resources and increase death."

Their understanding of different variables shows that they have fully grasped the model of tipping points and threshold dynamics. In the above quote, they accurately recognise that ICU strain is dynamic that varies according to the rate of illness spread, the number of patients who recover, and the way policies are put in place. Similar to this, they recognised in the marketing model that a product's tipping point is influenced by a variety of factors, including consumer behaviour, product quality, and outside interventions.

> When the product quality and customer loyalty reach a high-level point, there is a tipping point. This will ensure that all customers who try the product for the first time will become loyal and will try the product second time.

Another student from the experimental group states, "A tipping point occurs when the product quality becomes good enough (threshold), leading to increase customer satisfaction and repeat purchases."

In summary, the students from experimental group demonstrated their ability to apply the tipping point framework in both models by recognising these conceptual connections. Their responses show that they have a deeper and more adaptable understanding of how systems work, making it easier for them to think critically about how thresholds function in various real-world scenarios.

Surface-Level Features

Tipping points are characterised by their surface-level characteristics, which according to the findings of the study, consist of observable triggers, visible trends, and direct impacts that indicate when a system is rapidly approaching a critical threshold. During the study, both groups identified important aspects of exponential growth, sudden shifts, and external influences, based on the responses of both groups, this theme includes the following sub-themes:

1. **Observable Triggers of Tipping Points**: It discusses the clear signs that indicate when a system is approaching or has reached a tipping point
2. **Rapid Spread and Exponential Growth**: Recognition of how change accelerates quickly after the tipping point is crossed.
3. **Key Influencing Factors**: Awareness of specific input variables (e.g., contact rate, marketing exposure) that result in tipping points.

Observable Triggers of Tipping Points

The findings of the study indicate that both groups recognised sudden spikes and transitions as key indicators of tipping points in the malaria and marketing models. A student from the control group described the concept in simple terms, "If a large number of mosquitoes become infected with malaria, the spread of the disease may reach a tipping point and the cases of malaria will increases rapidly."

Although, this explanation correctly identifies a tipping point as a moment of rapid spread but most of the students were unable to explain why this occurs beyond the presence of infected mosquitoes. Similarly, the control group explains this concept in the marketing model, stating, "A tipping point can be seen when a product becomes popular to spread rapidly through a population. This means that small changes in the behaviour of consumers can lead to a rapid increase in product sales."

While their responses clearly capture an observable shift in consumer behaviour, it fails to take into consideration the external influences, interventions, and structural elements that contribute to these rises. In contrast to the control group, students in the experimental group provided more detailed insights into how different factors can play a role in influencing tipping points. One student explains, "The peak occurred at the beginning of the disease spread. As the disease started to infect more people, the community started to control the disease until returned to a stable state." This response goes beyond just defining the concept rather it delves into the complete system response, pointing out the change from instability to stability. Unlike the control group, which focused on the tipping point itself, the experimental group acknowledged the processes that follow, such as immunity development and policy interventions.

A similar depth of reasoning was noticed in their responses in marketing model, where a student notes: "There is a tipping point where the number of unaware people decreased because of advertisement campaigns. There is a dramatic increase of sale in the first weeks because of high adoption fraction and advertisement campaigns."

In this response, it can be seen that how the student was able to recognise external influences such as advertisements and seasonal trends which affect the tipping point, showing a more nuanced understanding of consumer behaviour dynamics than the control group.

Rapid Spread and Exponential Growth

The analysis of the study shows that both groups understood the concept of tipping points which are frequently associated with exponential change, such as the spread of malaria or the adoption of a product. A student from the control group highlights, "In malaria, tipping points happen when there is a dramatic increase in the number of deaths after increasing the number of mosquitoes contacted per day and the infection probability parameter."

This response reflects an understanding of how small changes in contact rate can cause infection rates to rise exponentially, which supports the idea of threshold dynamics. However, the experimental group, compared to the control group, provided a more analytical reasoning, discussing how these shifts could be influenced or delayed. One student explained: "A tipping point ca be seen when the product quality is good, the customer satisfaction and the purchase times increase." This response shows an understanding of how adjusting system variables can influence tipping points, highlighting how external interventions can actively shape the change.

In conclusion, the Control group characterised tipping points as observable moments of exponential change, but they did not investigate the reasons for their occurrence or the potential ways in which they could be influenced. The Experimental group investigated the factors that contribute to tipping points, their evolution over time, and the interventions that can either alter or postpone them. By examining the factors that influence the tipping point in addition to the tipping point itself, the experimental group exhibited a more comprehensive understanding of system dynamics, thereby employing a more critical perspective to the examination of both models.

Key Influencing Factors

The findings revealed that students from both the control and experimental group identified key variables that contributed significantly to the emergence of tipping points in both models. Their responses indicate that they have comprehended the concept of tipping points, which are characterised by the occurrence of specific critical conditions or values, rather than occurring at random. A control group student described how multiple factors could influence system change, "The tipping point is the threshold that changes from one situation to another because of some factors. For example, in the case of malaria the hospital capacity affects the numbers of ICU and deaths as the disease spreads fast."

This quote above suggests that the students were able to understand how external factors, such as hospital capacity, can directly affect whether a system goes into crisis. Similarly, while discussing the marketing model, another student in the control group accurately described how popular a product becomes as a tipping point, stating, "The tipping point in marketing model is seen when a product reaches a high level of acceptance from the consumers." This indicates an awareness that consumer adoption can initiate rapid changes in product success, despite the fact that the explanation remains centred on the outcome rather than the fundamental factors that contribute to it.

On the other hand, the students from the experimental group showed a more in depth understanding of the concept as their responses were focused on specific variables in the model. One student explains, "Tipping points happen if the infection period exceeds the timing, there is a high risk continuing spread which need to apply strong control measures." It indicates that he comprehended the significant impact of changes in specific factors, such as the duration of someone's infection, on the disease's transmission. Another student in the experimental group also linked the success of a product to a clear factor, saying, "Tipping points may occur when the adoption fraction (number of customers who buy the product) reaches a high level."

This shows they understood that even a small improvement in product quality could lead to a big rise in customer interest and sales. In general, the results suggest that, similar to session 1, both groups identified significant factors. However, the experimental group was more found to be more skilled in identifying specific variables and describing the impact of these variables on the model's behaviour which shows a clearer understanding of what causes tipping points in complex systems.

Session 3. This section summarises the results of the third comparison task, which required students to evaluate the structural similarities and differences between the Wolf-Sheep predation model and the COVID-19 model. According to the findings of the data, both models show complex systems that involve interdependent variables and dynamic equilibrium. However, they differ considerably in terms of domain (e.g., human versus ecological systems), complexity level, and response to interventions. The structural-level aspects which are discussed below reflect students' knowledge of causal mechanisms, system behaviour across time, and their understanding of how internal and external factors interact to determine system outcomes.

Structural-Level Features

Control Mechanisms

Within this sub-theme, students articulated how internal components of a system work together to produce emergent behaviours through feedback loops and reciprocal dependencies. The COVID-19 model, and the Wolf-Sheep model depend upon internal processes which control system evolution, including infection transmission through human interaction or changes in the predator-prey population cycle due to births, deaths, and resources.

Students from both groups demonstrated an awareness of these internal dynamics. A control group student observes, “Both systems show that parts of the population depend on each other (interdependency.” It demonstrates comprehension of the fact that the participants in each model, whether humans in the COVID-19 system, or wolves, sheep, and grass in the ecological model, work as a unit. Rather, their behaviours and conditions are actively interdependent on the other elements within the system. For example, the number of infected people in a population has a direct effect on the number of newly susceptible people, similar to what happens when the number of wolves increases; the sheep population decreases.

The students from the experimental group also expanded the same understanding: One student explains “In both models, the behaviour of individuals or sheep and wolves impacts the dynamics of the overall system.” It articulates understanding of how actions or behaviours of individuals (e.g. human contact, sheep grazing) are, in fact, actions that are summed up at a population level. It also captures the idea that even the most basic of interactions, when repeated over time, can result in system-wide complex behaviour.

What sets the experimental group’s response apart is closer affiliation with systems thinking. Their description implicitly advocates micro to macro linkage, which is the degree of change of individual constituent parts and how these changes can propel through the system to ultimately determine the course within which the systems evolve. This level of understanding reflects a deeper appreciation of complex systems and emergent phenomena. In the end, both groups agreed that the system is driven by value interactions, but students in the experimental group were more inclined to give answers regarding the relations of the interactions within the system models.

Model Complexity and System Dynamics

This sub-theme is focused on the challenges posed by an increased number of variables, decision-making agents, and rules in both systems. The COVID 19 model simulates disease transmission among humans while using infection rates, recovery rates, immunity, policy interventions, and healthcare resources as multiple variables. On the other hand, the Wolf-Sheep model uses fewer complex rules in simulating ecological processes in a predator-prey system, such as birth, death, and resource expenditure. These systems possess fundamental differences in design which determine their behaviour and responses to changes and stability over time.

Both groups of students learned to differentiate this level of complexity, but the students from the experimental group showed greater understanding of this dimension. One student observes, “Covid-19 model is more complex than the Wolf-Sheep model with some features. Covid-19 model has policies to control infection, but the other model does not have any policies to decrease the problem effect.” This shows students’ awareness that the COVID-19 model contains multiple layers of policy-based intervention, making it less predictable and more responsive to external inputs.

Another experimental group student reinforced this difference by stating that, “The Covid-19 model had much more comprehensive parameters.” It suggests that students not only noticed the number of adjustable inputs but also understood how these parameters shape system outcomes.

The students in the control group recognised the various model structures but mainly discussed system characteristics instead of computational complexity. One student noted, “The difference is that in the first model we are dealing with people’s lives, while the second one is dealing with sheep and wolves—the animal cycle.” This comparison between models appears more basic as students framed the differences using domain contexts instead of model intricacies. Participants from both groups recognised that the Wolf-Sheep model was simpler than the COVID-19 model. However, students in the experimental group provided more precise and system-focused explanations, suggesting an in-depth understanding of internal rules and variables that influence model behaviours.

Resource Dependency and Population Balance

This sub-theme demonstrates how essential resource availability levels as well as

resource exhaustion patterns affect population survival together with system results in both models. Medical system capacity functions as a decisive factor to determine if health services can properly handle an outbreak in the COVID-19 model. Population survival in the Wolf- Sheep model depends on available ecological resources which provide grass to sheep and prey to the wolves. In both cases, when essential resources reach depletion, populations decline or collapse, demonstrating the systems' sensitivity to resource constraints. The students in the control group investigated system limitations through their discussions. One student noted, "Both systems need to ensure that there is enough resources available to live like hospital services (COVID-19) and sheep (wolf). If those resources decrease, it leads to death."

The experimental group offered a similar insight but added context about system stability: "The major difference is the Wolf-Sheep model is found in an ecosystem and is difficult to change the dynamic, while the COVID-19 model faces many interventions to avoid increasing death." The students demonstrated an advanced understanding about managing human systems through resource management whereas ecological systems remain restricted to non-intervention approaches. Another experimental group student remarked, "We should not go beyond the balance (equilibrium), because this will cause outbreak and death," highlighting how balance is maintained by allocating resources correctly, and how even small changes can cause the whole system to fail.

The data reveals that both student groups understood that system survival depends on resource availability for stability. However, students in the experimental group noticed how resource availability affected comprehensive system operations and interventions at critical thresholds thus exhibiting more advanced understanding about system modification.

Equilibrium and Stability Mechanisms

This sub-theme focuses on how self-regulating processes and internal feedback loops

help systems either remain stable or become unstable over time. The COVID-19 and

Wolf- Sheep models both exhibit dynamic equilibrium, which is a condition where system variables change but stay within a manageable range because of balancing factors. In the COVID-19 model, this could be observed in the fluctuations in infection rates brought on by immunity, recovery, or intervention measures, while in the Wolf-Sheep model, equilibrium is reached when the numbers of predators and prey fluctuate about constant values, supported by cycles of birth, mortality, and food availability.

Students from both groups shared their views that how equilibrium works within each model. A control group student explained, “They both show dynamic equilibrium. There are changes at micro-level, but the overall system remains the same.” Another control student illustrated this with a more specific comparison: "The main similarity between the two systems is how different factors affect equilibrium. In COVID-19, the changes of patients when recovering and dying can show the equilibrium. But in the Wolf-Sheep model, there is balance between different species." This response demonstrates an ability to compare how equilibrium is maintained in different domains—public health and ecology—through cyclical adjustments in population states.

Experimental group students shared the same views and understanding but added more detail about the nature of these cycles. One student says,

> Initially, both systems involve equilibrium. In the COVID-19 model, increase in infections is followed by a decline. The Wolf-Sheep model shows an increase in prey leads to an increase in predators, then causing a decline in sheep, followed by a decline in wolves.

This response shows students understand how population dynamics work in both systems and how factors that depend on each other change over time. Another student from the experimental group expanded further: “Both models work on similar concept which is dynamic equilibrium which means two different parts of the same system are related, so if one part change, the other part will change.” This explanation shows a better understanding of how opposing forces, such as infection and recovery or predation and reproduction, keep things stable by always changing.

In summary, the findings of the study indicate that students in both groups possessed a sufficient comprehension of equilibrium in complex systems. However, the participants in the experimental group demonstrated improved abilities to articulate the patterns in which systemic cycles persist due to their improved comprehension of the dynamics of system balance in ecological and epidemiological frameworks.

Surface-Level Features

This theme exposes the direct observations of students regarding the visual aspects of models, including their representations and environmental design, along with the behavioural display variations.

Graphical and Visual Representation

In this sub-theme, the students from both groups discussed how the models were represented visually. However, the responses from the experimental group were limited compared to the control group. The experimental group simply state “Both show graphical representation.” A similar remark was made by another student in the same group, repeating the exact phrase. These responses indicate that students acknowledged the use of simulations in both models but did not elaborate on the differences in information visualisation.

In contrast, students in the control group made more in-depth observations of the visual and spatial features. A student explains “In the epidemics system, the environments for the population show interesting dynamics. In the wolf-sheep model, the population moves freely.” Another student highlighted the spatial configuration more clearly, "The Wolf-Sheep model allows movement in 2D space that makes individuals look separate, whereas the COVID-19 model work on the probability of whether people around you affected you or not."

These responses from the control group show a more sophisticated comprehension of how the models simulate interaction. In Wolf-Sheep, this is accomplished through physical movement and visual closeness, while in COVID-19, it is accomplished through predictive infection. Although both groups agreed that images were present, the control group gave more structured responses to how these visuals connected to system dynamics.

Spatial and Environmental Configuration

This sub-theme examines how each model's environment influences agent interactions and system dynamics. Students in the control group examined how space and external influences form models differently. One student remarked: “The wolf-sheep model doesn't consider any external factors that might affect its dynamics from inside,” indicating that the Wolf-Sheep model works in a closed biological loop, where most of the interactions happen in the environment and are controlled by natural processes.

Another control group student highlighted the role of spatial configuration, stating: "The Wolf-Sheep model allows movement in 2D space that makes individuals look separate, whereas the COVID-19 model work on the probability of whether people around you affected you or not." The quote shows the students’ understanding of how the Wolf-Sheep model relies on physical positioning and movement within a space, whereas the COVID-19 model is driven by probabilistic interactions based on proximity, regardless of the spatial direction.

In conclusion, while both groups recognised visual representation, the control group provided greater descriptive insights into spatial structure and its relationship to system operation. The experimental group noted visuals but did not go into detail, indicating a surface-level assessment of the models' graphical elements.

Nature of Entities and Domain Context

This sub-theme examines the students' understanding of the basic elements or agents in each model, comparing people in the COVID-19 model to animals in the Wolf-Sheep model, and how this affects their interpretation of model behaviour. The findings of the study indicate that students from both groups clearly distinguished the models based on the types of entities involved. A student in the control group points out, "The COVID-19 model is a human disease model, while the Wolf-Sheep model involves animal species." Similarly, a student from the experimental group noted: "The difference is that in the first model we are dealing with people's lives, while the second one is dealing with sheep and wolves the animal cycle." This quote underscores a more profound emotional or ethical connection with the COVID-19 model, as it emphasises the importance of public health and human existence, in contrast to the ecological focus of the predator-prey model. While both groups displayed this fundamental difference, the experimental group was observed to have more understanding about the model and its implications; for example, how public health interventions are relevant to the COVID-19 model but not normally implemented in ecological systems such as the Wolf-Sheep model.

RQ4: Qualitative Results from Self-Reports and Focus Group Interviews

As explained in Chapter Three, a qualitative analysis of participants' self-reports (gathered at the end of each content session) and their contributions to the focus-group interviews (held after all instructional sessions and testing had concluded) yielded four recurring themes that were addressed in the feedback from the participants:

- Engagement and Positive Learning Experiences,
- Challenges or Barriers to Effective Learning,
- Comparison of Learning Methods, and

- Suggestions for Improvement.

The perspectives of the students on these four topics are summarised below in turn.

Theme 1: Engagement and Positive Learning Experiences

One of the advantages of using Agent-Based Models (ABMs) in epidemiology instruction is their ability to engage students through interactive, visual, and exploratory learning experiences. Compared to traditional teaching methods, ABMs allow students to manipulate variables, observe dynamic system behaviours, and apply their learned knowledge on real-world scenarios. However, the degree and nature of engagement varied significantly between the Productive Failure (PF) as an experimental group and Direct Instruction (DI) as a control group. The PF group, which encouraged self-directed exploration prior to instruction, often exhibited higher engagement levels while the DI group, which received the instruction first followed by practice tasks, presented engagement levels that were more dependent on the clarity and structure of teaching content.

Students in both groups highlighted aspects of ABM-based learning that enhanced their engagement, including the clarity of content, the role of visual aids, opportunities for hands-on experimentation, the flexibility of parameter adjustments, and the ease of using the learning platform. The following sections explore these factors in detail.

Clarity and Comprehensiveness of Content. The way instructional content was presented played the primary role in shaping student engagement and their learning experience. Students demonstrated greater confidence in their knowledge and participated more actively when concepts were explained clearly and presented in a logical order alongside structured learning activities. The DI (Control) group received major benefits from structured explanations that simplified complex epidemiological concepts while providing a systematic learning approach. Students found that the organised and sequential delivery of material was essential for their active engagement. A student expressed appreciation for how the lessons followed a clear structure; "Actually, the material was smooth, easy to absorb and understand. The sequence of activities simplified the content well enough. Applying the knowledge on the simulator improved the understanding even more." Similarly, another DI student praised the way complex systems were simplified through instructional materials. According to the student, "It was an impressive attempt to simplify the problem of chaotic systems (they appear chaotic). The videos were short enough not to be daunting."

Additionally, DI students also valued the clear presentation of theoretical content, as noted by one participant, "Organized, detailed lectures made everything work well." Other students also shared the same sentiment, showing that the systematic organisation of lessons helped them stay engaged and focused. According to them, the work was well-arranged and clear, and the concepts were explained in a good and systematic way.

PF (Experimental) group experienced content clarity differently because they followed an exploratory learning process that was not structured. Students who initially struggled with understanding needed time to adjust to this learning style but eventually found it effective after adaptation. A student from PF reported that they faced difficulties due to the initial absence of structure. According to the student, "this type of learning started to work for me. And it is easy to get these concepts."

However, PF students started to value the clear understanding that developed from their discovery-based learning as they adapted to the instructional approach. A student discovered that their understanding improved when they studied relevant case studies and models. The students expressed, "the way the spread and malaria disease complication were easily explained by the model made it much clearer." Some of the students perceived the content and learning structure as effective, even though it was initially challenging. According to them, the content and length of videos were perfect, and the concepts were very clear and straightforward to understand. One of the students said, "I have run more smoothly through the tasks after getting used to the teaching method."

Several PF students indicated that the course delivered a comprehensive theoretical knowledge base which led them to develop a strong understanding after overcoming initial learning challenges. They observed that the model had been clearly described. One of the PF students mentioned, "approximately everything works fine, starting from the technical stuff ending with the last activity in the model." For some, the structured explanations provided later in the PF sequence helped consolidate their learning, particularly when reinforced with real-world examples. According to a student, "the doctor's explanation of the material, along with some stories about cases, helped as well."

In conclusion, DI students found structured explanations beneficial because they gained immediate understanding, which helped them learn concepts without feeling overwhelmed. However, some students experienced limited deeper engagement because they received information in a delivered format rather than through active discovery.

In contrast, PF students faced additional initial difficulties because they needed to find patterns by themselves. On the other hand, the students from the DI group who became accustomed to the method reached a deeper understanding of the material, which allowed them to interact with it more effectively.

Enhancing Understanding Through Visual Aids. Visual aids, especially dynamic models, graphs, and instructional videos, significantly enhanced student engagement and conceptual understanding in both the PF and DI conditions. Students developed a better understanding of complex epidemiological concepts by visualising interactions between variables alongside real-time parameter manipulation and disease tracking rather than through conventional textbook learning.

Both groups of students discovered ABMs to be highly engaging because they enabled visualisation of complex relationships beyond what traditional lectures and readings could achieve. Students valued the quick feedback they received from testing different scenarios which enabled them to grasp epidemiological systems in a more intuitive way. A PF student explained that the models offered a detailed comprehension of epidemic dynamics which proved difficult to achieve through conventional teaching methods.

> For me, I think the modules are very helpful. They give me a very deep idea about epidemics, like how it has different parameters that I can change. It also has a map and other graphs that show me the change. This gave me a very deep understanding of epidemics. In the traditional way, I don't get this depth because there are no graphs.

Similarly, another PF student emphasised how the ability to manipulate visual models helped clarify complex interactions. According to the student, "the visual presentation makes it easy to imagine the scenario, and the extent and rate of spread."

In contrast, DI students typically used visual models as tools to strengthen instructional explanations. A DI student remarked that visual examples made abstract concepts easier to understand.

> It was very good in clarifying the whole image about every example that you gave to us. So the best thing about it was the clarification of whatever you can give us as an example. It was easier to understand and to comprehend whatever, that could, be explained to us.

A second DI student supported this viewpoint by highlighting that those visual tools transformed learning into an intuitive and approachable process.

> Whenever you see something in front of your eyes, on the laptop or on the screen, as, you know, as a life example in front of us, it will make anything easier to understand. So the thing that I like most about this is how easy everything went for us to understand everything.

In conclusion, the use of visual models was vital for engagement and comprehension across both groups yet manifested through different methods. PF students used visual models as their main learning resources whereas DI students utilised them to supplement instructional explanations.

Both PF and DI students found instructional videos extremely helpful because they offered structured explanations of complex subjects in a concise format. Students learned material more effectively through short videos because they provided organised content and clear examples to reinforce learning. One DI student shared that the instructional videos successfully reinforced key concepts and maintained an engaging yet manageable learning experience. The student said, "the voice in videos is clear and good. The videos were short and nice, making them easy to follow."

Similarly, one of the DI students shared that animated videos helped improve learning outcomes especially for medical students who depend on visual learning tools. According to the student, "for medical students, I noticed that we like to see more animated videos and examples to comprehend and understand topics. This is a very big deal and an important step to take."

Videos played an essential role in concept clarification for students in the PF group after their initial interaction with models. A student in the PF program explained that videos assisted them to understand concepts which were hard to comprehend only through experimental methods. The videos explained very well and removed all the missing gaps that the students had in the modules before. Another student appreciated the clarity of video explanations, stating, "the videos are informative and help clarify some aspects of the graph." Interactive models alongside explanatory videos functioned as essential tools for comprehension and engagement during both PF and DI instructional conditions. Students who preferred independent exploration in the PF condition found that visual models provided opportunities for content interaction which helped them enhance their critical thinking skills. DI students found visual aids beneficial because they helped support the structure of lecture explanations, whereas some of the students from PF group also mentioned experiencing challenges to understand system behaviours due to insufficient guidance.

Exploratory Learning and Experimentation. ABM-based learning featured exploratory learning which enabled students to test hypotheses actively while manipulating variables and performing trial-and-error experiments. Traditional lecture-based instruction delivers knowledge passively while ABMs promoted active and dynamic interaction with epidemiological concepts. The exploration methods between PF and DI groups showed distinct differences. PF students practiced open-ended experimentation that required them to address challenges before receiving explanations from instructors. Students developed critical thinking and engaged deeply through this approach yet experienced initial confusion. Structured sequences guided DI students through learning which limited discovery opportunities yet enabled faster acquisition of essential knowledge.

The PF group gained significant learning benefits through exploratory learning because students could test various hypotheses without risk which enhanced their knowledge of epidemiological systems. This method engaged numerous students because they could directly witness how their choices led to immediate outcomes.

Through scenario exploration, a PF student managed to grasp system dynamics more effectively.

> I liked the interactive learning method because we can simulate different scenarios and observe the outcomes in real time. This deepened our understanding. Also, I liked how this system enabled us to visualize complex systems by understanding the interactions between agents. And this is exactly what emergence means.

Another PF student appreciated the ability to experiment without real-world consequences, she states, "We can try multiple hypotheses according to adjusted rules and then see the outcomes without real-world consequences."

However, as compared to PF students, DI students had fewer opportunities for open-ended exploration, as their learning was guided by structured instruction. Some of them expressed a desire for more hands-on engagement, as one student noted:

> I think the best for me is to have the subject and then going deep on it by myself, this usually works for me. On the other hand, I think it is useful to have changes in the ways how do we get the information. I almost agree with my colleague. When I watched the videos and learned how to use and change the parameters to see the results, I then started to explore the parameters easily.

The exploratory learning approach greatly improved students' development of critical thinking abilities. PF students discovered that their analytical and reflective thinking skills improved through independent model interpretation.

One learner explained that responding to questions before video instruction contributed to their critical thinking development. According to the student, "I think answering questions before watching the videos can improve our critical thinking and prediction because we have to think about all the possible things before we know the exact things."

Another PF student described how ABMs provided an opportunity for self-guided learning, which they found beneficial:

> The best for me is to have the subject and then go deep on it by myself. This usually works for me. On the other hand, I think it is useful to have changes in the ways we get information.

Furthermore, exploratory learning enhanced student comprehension while also increasing their enjoyment and motivation especially with PF students. The process of pattern discovery and connection formation provided an engaging and rewarding experience for numerous participants. According to one student, learning became more pleasurable when actively exploring models rather than using traditional teaching methods. Another student appreciated the hands-on approach of ABMs, which encouraged repeated experimentation and engagement, "I enjoyed exploring the models by trying and analysing each time."

Students in DI classes received organised instructor-led training but were not encouraged to develop independent critical thinking skills. Students underwent organised educational procedures but found few opportunities to question their pre-existing beliefs or examine different hypotheses without guidance. Self-directed exploration enabled PF students to enhance their critical thinking abilities while DI students' analytical skills were limited by structured instructional methods that restricted independent topic analysis.

Theme 2: Challenges in Learning with ABMs: A Comparison of PF and DI Approaches

While ABMs allowed researchers to examine complex epidemiological systems interactively, students in both PF and DI groups encountered multiple challenges. The inherent complexity of ABMs created barriers for students who also lacked structured guidance in PF and faced limited model flexibility along with technical problems and unclear question formats. Initial cognitive load during exploration led to deeper content understanding for PF group students compared to those who received direct explanations first. Through structured guidance students in the DI group followed a straightforward learning path but had limited opportunities for deep exploration.

Participants assigned to the Productive Failure (Experimental) and Direct Instruction (Control) conditions experienced multiple educational challenges but also gained from the interactive nature of Agent-Based Models (ABMs). Students' learning difficulties revealed five primary sub-themes, as identified through the thematic analysis conducted in this study.

1. Complexity and Confusion
2. Lack of Instructor Guidance
3. Limited Flexibility
4. Technical Difficulties
5. Unclear Question Formats

Complexity and Confusion. The ABMs required students to manipulate numerous parameters while interpreting data trends and system behaviours but presented significant challenges for PF students who did not receive direct instruction prior to model usage. The task of connecting numerous variables while evaluating their relationships led students to experience cognitive overload. A PF student described this frustration as "When the model grows in complexity, there are too many variables to manipulate. It feels daunting to go through each variable again and again."

Similarly, another PF student highlighted the difficulty of tracking relationships between multiple parameters, emphasising the cognitive burden associated with model-based learning:

> I found it difficult to manipulate the first variable and then go and manipulate 10 variables next. Then I would like to observe what is the relation of the 10th variable to the first variable. So, I go back to manipulate the first variable again.

Another student noted that navigating through the different system components without prior instruction felt overwhelming stating, "A new experience, things seemed not connected at first, and it took time to understand the relations and principles."

In contrast, DI students found ABMs easier to understand because they learned through a structured introduction process. Several students had trouble interpreting graphical representations such as charts and graphs which were demonstrated through an individual student's challenging experience, "The graphs are sometimes difficult to interpret."

The absence of immediate instructional support led to higher cognitive overload for the PF group which prolonged their conceptual understanding. On the other hand, DI students benefited from clear explanations that lessened their confusion but limited their ability to pursue deeper understanding on their own. A learning approach combining guided instruction with self-exploration opportunities may help solve these educational challenges.

Lack of Instructor Guidance. The results indicate that the PF students faced uncertainty and frustration because they received no explanatory guidance before using the models. Students had trouble interpreting models on their own which led them to ask for more structured teaching assistance. According to one of the students, "The concept of the tipping point related to the model needs further explanation."

Others emphasised the need for clearer introductory explanations before engaging with the ABMs because according to a student, "Giving more clear ideas at the beginning about these models was necessary for me."

Furthermore, some PF students pointed out that the lack of immediate feedback on their interactions with the models hindered their learning experience. One of them shared, "One limitation was that we didn’t receive any feedback on our answers."

In contrast, DI students received instructional guidance at the beginning which helped them establish a solid conceptual base before they started working with models. According to one of the DI students:

> The thing is that, what I think it was really helpful because when we learn about the concept of epidemics, even if we took some details, it is mostly comprehensive, and we could sometimes miss some points. You know? And as we learned in the complex system that there were, like, many factors or the micro factors that could affect the whole thing. So it was helping us to focus not only on the big ideas, but also on the details behind the whole fact and not to miss some point.

After these instructional initiatives, a number of students continued to experience uncertainty which led them to ask for more comprehensive explanations about challenging epidemiological subjects; for example, "It was very clear for the other examples and for the other concepts that you gave us, like tipping points, because I couldn't understand them at first before you gave us the videos and the explanation."

In conclusion, the PF group needed structured guidance during their early learning stages whereas DI students experienced directed teaching but had few opportunities to explore topics independently. Incremental guidance that merges exploratory tasks with timely explanations produces superior student understanding and active participation.

Limited Flexibility in Parameter Manipulation. Through ABMs, students can study different parameters to explore diverse epidemiological outcomes. While learners appreciated the feature, they observed that excessive flexibility led to unrealistic or impractical scenarios. A PF student expressed concern about how unlimited parameter changes could produce unrealistic results, "The ability to change parameters is nice, but it can create unrealistic situations that affect the validity of the model." He further stated that it could undermine the model's credibility.

According to another PF student, this level of control does not match real-world scenarios which creates problems for deriving practical conclusions. According to him, "We have many parameters, which is nice to change the situation of the model as you want. But at the same time, it creates unrealistic situations."

DI students experienced the structured method as limiting because it stopped them from investigating different epidemiological scenarios. One of them shared, "Actually, it's good if the information stay until the exam. The traditional learning mainly depend on memorization. it doesn't allow us to, let me say, to think outside of the box."

In conclusion, the PF condition provided enhanced flexibility but caused unstructured exploration which hindered the process of deriving meaningful conclusions. DI students experienced intensive learning but lacked opportunities for independent experimentation. When students work with structured parameter manipulation and moderate flexibility, they exhibit improved learning outcomes and increased engagement.

Technical Difficulties. The learning process for both conditions faced disruption due to technical problems such as slow system responsiveness and system crashes. PF students experienced greater disruptions from technical problems because their learning depended more on ABMs. One PF student reported, "The program crashed while I was answering the last quiz, so I had to use the information from the videos instead." While another one highlighted slow system responses, which interfered with model interactions, "The speed of responding of computer-based models was slow." In contrast to this, DI students also experienced technical limitations, though these were less disruptive due to the availability of alternative instructional materials as they also found, "The website (AnyLogic Cloud) was slightly slow."

PF students faced greater obstacles from technical issues because they depended on ABMs for learning independently, but DI students managed to work through instructor-led explanations even when the system slowed. Technical reliability becomes essential for keeping students engaged when they participate in exploratory learning activities.

Unclear Question Formats. The findings indicate that both groups faced issues with unclear assessment questions, but PF students had greater difficulty because their understanding of assessment structure was compromised without direct explanations.
One PF student commented on the difficulty they faced when trying to comprehend assessment questions. He shared, "I did not understand the questions well." Another student noted that some questions were excessively long and complex: "The question has a word limit of 350, which is long and hard to reach, plus the time limit is short for two questions of 350 words each in 45 minutes only." Even DI students found some assessment questions unclear, as one remarked, "For most of the questions, I wasn't quite sure what was required of me."

In conclusion, PF students required better-structured questions since they received less guidance about the answer context compared to DI students who required clearer question formats. Clearer questions and assessments that match instructional strategies enhance student performance while lowering frustration levels.

Theme 3: Comparison of Learning Methods

This theme is divided into three sub-themes which are: (1) the enhanced learning opportunities provided by ABMs, (2) the limited flexibility of traditional learning approaches, and (3) the practical applications of ABMs in real-world learning scenarios.

Enhanced Learning Through ABMs. Through ABMs students actively learned complex epidemiological concepts which traditional methods struggle to teach. ABMs empowered students to control variables and conduct hypothesis tests while observing live interactions whereas in traditional learning students only memorised content from lectures and books. The PF group students discovered that ABMs facilitated deep learning by promoting systems thinking while enhancing problem-solving abilities and ensuring long-term concept retention. A PF student reported that their knowledge consolidation improved with ABMs compared to traditional lecture methods: "Productive failure has done a better job in terms of consolidating the information and recallability afterwards. I can immediately recall the information even after a long time, which doesn't happen with traditional lectures."

Another student highlighted how PF forced them to think critically and apply their learning across different contexts, enhancing knowledge transfer:

> Protective failure mainly targets to what extent we can transfer knowledge to different contexts. In traditional teaching, we might hear lots of information within two hours, but later, we forget most of it because there's no problem-solving, no brainstorming, and no animation or graphics to engage us.

DI students acknowledged the effectiveness of ABMs as educational tools but mainly applied them to reinforce knowledge rather than to explore new information. The DI students found ABMs beneficial because they helped demystify abstract epidemiological concepts. According to a student, "The complex system clarifies the concept of epidemics. For example, in complex systems, one small thing changes the whole system. In epidemics, if one person doesn't wear a mask, this could increase ICU cases in the future."

Similarly, another student in the DI group recognised that ABMs helped visualise relationships more effectively, stating, "It was very helpful because I could see what was happening in the system, imagine it, and even be involved in manipulating it. That gave me a better understanding of different relationships."

Limited Flexibility in Traditional Learning. The findings of the data indicated that traditional learning methods failed because their rigid structure emphasised memorisation above practical application. Students belonging to the DI group found traditional learning methods useful for retaining information in the short term but lacking in both flexibility and practical application in real-world scenarios.

The DI student pointed out that traditional education gives more importance to rote learning than it does to critical thinking. He stated, "Traditional learning mainly depends on memorisation. It doesn't allow us to think outside the box."

Another student reinforced this point, explaining how traditional education often fails to encourage knowledge application, "We always have difficulties applying what we learn in college. Most of the time, it feels like we're just memorising things. If a question requires application rather than just recalling facts, it becomes much harder to answer."

A significant limitation of traditional learning was also evident in exams, where students struggled to apply knowledge to novel problems. According to one of the students, "Even when we understand the concepts, exams are difficult because they ask practical questions that we haven't been trained to apply."

PF students appreciated the ABM-based approach since it enabled them to execute multiple scenarios while learning through experimentation. According to a student in the PF program, ABMs produced a richer interactive learning experience compared to traditional lectures, "Traditional learning gives information linearly. You learn concepts, see case studies, and move on. But with ABMs, I was able to actively engage with the material, experiment, and see relationships unfold in real time."

Another student emphasised how ABMs facilitated a more intuitive understanding of epidemics. According to the student, "systems thinking was incredibly helpful in understanding epidemics. Seeing how individual behaviours influence overall disease spread was much clearer when modelled in a system rather than simply reading about it."

According to the findings, although traditional education methods delivered structured knowledge acquisition but failed to provide the flexibility required for students to achieve deeper conceptual understanding and practical application. In contrast, the exploratory nature of ABMs helped PF students develop systems-thinking skills and allowed them to engage more meaningfully with the material.

Practical Application of ABMs in Learning. ABMs demonstrated considerable value by linking theoretical understanding to practical applications. Students from PF and DI groups recognised ABMs as valuable tools for epidemiological analysis but applied these models in varying ways. According to the analysis of the data, PF students recognised ABMs as extremely effective for disseminating knowledge across multiple academic fields beyond epidemiology. ABMs sparked new ideas for the student about how to use these models outside healthcare applications. One student highlighted how ABMs encouraged them to think about broader applications beyond the health field, stating "ABMs helped us apply concepts not only in health and medical issues but also in different fields, like environmental science and disaster management."

Another PF student emphasised how ABMs helped them understand complex, emergent behaviours that might not be observable in real-world experiments, "For example, emergent behaviour—if it were a real experiment, we might not get to fully understand or observe it. But with computational simulation, we can visualize it clearly."

DI students valued ABMs for theoretical backing but felt that traditional methods were insufficient for applying the concepts in practical contexts. Traditional teaching methods provided students with essential conceptual knowledge without demonstrating practical real-world uses. According to a student from DI, ABMs improved their ability to conduct analyses in practical epidemiological situations. He says, "the model example was a very good way to demonstrate the idea discussed in each session. It made learning more practical and enjoyable."

A DI student observed that traditional teaching methods helped students understand theoretical concepts during lectures but did not support practical application of that knowledge. He shared "The slides and regular theoretical sessions work well for me in understanding the concepts in general, but not very well in applying them and knowing the effects of parameters."

In conclusion, PF students identified valuable practical advantages in ABMs due to their capacity to enhance knowledge sharing among various domains. In contrast, DI students appreciated ABMs because of their theoretical foundation and felt traditional methods failed to meet practical requirements for real-world situations.

Theme 4: Suggestions for Improvement

While ABM-based learning provided significant benefits in engagement, conceptual understanding, and real-world application, students from both the PF (Experimental) and DI (Control) groups highlighted areas for improvement. Six sub-themes are created based on their suggestions, which are discussed below.

1. Expanding the application of ABMs to other fields
2. Improving content depth to enhance comprehension

3. Integrating technology to enhance traditional learning
4. Providing clearer guidance and instructional support
5. Encouraging discussion-based learning for deeper engagement
6. Simplifying assessment formats to improve clarity

The following sections explore these areas in detail, integrating student feedback to provide a comprehensive analysis of potential improvements.

Application of ABMs in Other Fields. Many students recognised the potential of ABMs beyond epidemiology, which included social science research, chronic disease modelling, and healthcare administration. One DI student explained that ABMs are valuable for chronic disease modelling because they enable students to observe the effect of lifestyle factors on disease advancement. The student said, "In my point of view, ABMs can be helpful because they simplify the relation between simple agent interactions and the resulting complex effects. Maybe we can use them for chronic diseases, where multiple factors contribute to outcomes."

A student suggested introducing ABMs earlier in education: "I think you could implement this in schools before universities. You cannot just enforce this way of learning on people who have been using traditional methods for 18 years and then suddenly change the way they learn."

In conclusion, students from both PF and DI groups identified potential uses of ABMs across different disciplines. DI students investigated medical applications and PF students identified interdisciplinary uses. The relevance and impact of ABMs would rise if they were applied outside epidemiology.

Improving Content Depth. The findings of the study indicate that PF group students experienced hands-on exploration through ABMs but believed certain theoretical concepts needed better coverage. Students recommended supplementing the curriculum with detailed explanations and case studies alongside practical examples to improve learning reinforcement.

A student from the PF group expressed a desire for increased complexity in epidemic modelling, stating, "If more factors can be added regarding epidemics and how they might affect each other, it would be good." Another PF student suggested incorporating real-life applications to strengthen understanding: "Providing simulators for more specific issues on real populations would make the learning experience more practical." Students who received structured learning in the DI program also requested more theoretical explanations to help them better understand the models. As one student noted, "Explaining issues in theoretical courses in the context of complex systems will improve concept understanding."

PF students requested more theoretical guidance for their exploratory learning exercises while DI students needed practical examples to better grasp concepts. An educational approach that merges theoretical concepts with practical learning experiences could meet both needs.

Integrating Technology to Enhance Traditional Learning. Both PF and DI study groups recommended integrating more digital resources and simulation tools into regular lectures. The interactive nature of ABMs enabled students to move away from passive learning methods which they appreciated. During COVID-19, a DI student identified technology-based learning as valuable and advocated for its continued use in medical education, stating, "During COVID, doctors sent us lab simulators, and we applied them. It was great to learn that way. These kinds of technologies should be part of our curriculum permanently." Similarly, PF students wanted even more technological integration, with one student suggesting, "We should use more interactive tools like educational videos, additional case studies, and models to explore disease spread and intervention strategies."

Both PF and DI students valued technology, but PF students wanted more exploratory tools, while DI students saw technology as a reinforcement tool for structured learning. Integrating digital tools more effectively could improve engagement and understanding**.**

Need for Clearer Guidance. PF students struggled with ABMs because they did not receive immediate guidance, which frequently caused frustration. Students recommend implementing step-by-step tutorials along with guided walkthroughs and pre-recorded instructional content. One PF student suggested, "The program needs more help and explanation. Maybe provide pre-recorded materials or tutorial videos." Another PF student requested additional guidance on interpreting model outputs, stating, "It would be better if we were given sets of parameters to focus on while using the models and told what patterns to look for." In contrast, DI students generally felt they had sufficient guidance, though some still wanted clearer explanations in certain areas. As one student noted, "The models and videos were good, but some concepts needed additional clarification, particularly emergent properties in infectious diseases."

PF students needed more structured support to navigate their exploration, while DI students sought more clarity in theoretical explanations. Optional guided tutorials and structured walkthroughs could address both concerns.

Preference for Discussion-Based Learning. Many students emphasised the importance of discussions and collaborative learning, suggesting that peer interactions and instructor-led discussions could reinforce understanding and improve engagement. A DI student noted their preference for discussion-based learning, explaining, "I prefer to discuss more with the doctor, which makes me question myself and my knowledge." Another DI student added, "It’s easier to remember things when you’re asked about them and have to explain them, rather than just memorising from a lecture." Similarly, PF students saw discussions as valuable, particularly in clarifying uncertainties that arose during independent exploration. As one PF student expressed, "Live teacher guidance would produce better outcomes. Discussion sessions could help us understand things we missed during exploration."

Both groups valued discussions, but for different reasons—PF students sought discussions to clarify their independent exploration, while DI students saw discussions as a way to reinforce structured learning. Blending both approaches could enhance engagement and retention.

Simplifying Question Formats. A common challenge across both groups was the complexity of assessment questions. Many students found the wording difficult to interpret, making it hard to determine what was expected in their responses. One PF student recommended the use of simpler formats, stating, "The tests would be better if they included multiple-choice questions rather than long, complex responses." Similarly, a DI student emphasised the need for clearer language, noting, "The question structure is complicated, and some of the wording makes it hard to understand what's being asked." Overall, while PF students preferred simpler formats, DI students sought clearer question design. Refining assessment structure could help improve student performance and reduce cognitive load during evaluations.

Chapter Five: Discussion

In this chapter, the results of the study as they pertain to each research question (RQ) are discussed and related to the broader literature. The theoretical and practical implications of this research are also considered.

Discussion of Answers to Research Questions

PF Versus DI for Declarative and Explanatory Knowledge

RQ1: Does the PF condition lead to superior learning outcomes in *declarative* knowledge of epidemics, *declarative* knowledge of complex systems concepts in epidemiology, and *explanatory* knowledge of complex systems concepts in epidemiology, as compared to the DI condition?

As reported in Chapter Four, the post-test scores of students in the Productive Failure (PF) experimental condition were not significantly different from those of the members of the control Direct Instruction (DI) control group on any of the three sets of questions corresponding to the three categories of knowledge in view, namely:

1. declarative knowledge of epidemics,
2. declarative knowledge of complex systems concepts in epidemiology, and
3. explanatory knowledge of complex systems concepts in epidemiology.

This indicates that PF intervention had no significantly more marked impact on the students' development of these knowledge types than the DI approach had. In regard to the first two (declarative) types of knowledge, these findings align with prior studies demonstrating that declarative knowledge can be acquired without deep understanding as long as the content is delivered and reinforced in a structured manner. DI meets these criteria, whereas PF carries greater potential for the development of a higher-order understanding of conceptual knowledge (Jacobson et al., 2017; Kapur, 2014; Loibl & Rummel, 2014a, 2014b). These same studies, however, make it surprising that the second hypothesis (H2) of the study was not confirmed: As noted above, no significant difference was found in the present study between the PF and DI groups' post-test scores on explanatory knowledge of complex systems concepts in epidemiology. Although a few PF studies have reported similar results to those of the present study in regard to the acquisition of explanatory knowledge (Cao et al., 2024; Loibl & Rummel, 2014b; Mazziotti et al., 2019), the majority of PF studies in the literature have found PF to offer a clear advantage over DI where explanatory knowledge is concerned.

One possible explanation for the present study's failure to confirm H2 is the impact of including Agent-Based Models (ABMs) in the instructional/research design of both the experimental and control groups. That is, as Goldstone and Wilensky (2008) recognised and numerous others have demonstrated (e.g., Sengupta & Wilensky, 2009; Holbert & Wilensky, 2014; Silverman et al., 2021), ABMs constitute a powerful learning tool for promoting causal reasoning skills, the comprehension of mechanisms underlying phenomena, and the relation between the actions of individual agents and the properties of systems at the aggregate level (Levy & Wilensky, 2008). The learners in the DI condition no less than those in the PF condition had the opportunity when solving the challenge problems to explore different parameters in each model, observe the resulting changes, and learn about the targeted concepts based on their trials. The benefits of this active, student-centred learning with the ABMs may have been the overriding factor when it came to the development of explanatory knowledge in this study—dwarfing whatever benefit the PF condition, considered in isolation, afforded the members of the experimental group in that regard.

Some confirmation of this explanation comes from the qualitative data gathered from the self-reports and focus group interviews. Recall that multiple members from both the experimental and control groups commented positively on the level of interactivity and engagement that the ABMs introduced into the lessons. These comments suggest that the inclusion of ABMs in the instructional design played a major positive role in the subjects' learning process—to the point that it may have more or less levelled the playing field between the experimental and conditional groups in regard to the learning of explanatory knowledge.

One might object to this reasoning by pointing out that Jacobson et al. (2017), in a study of the development of declarative and explanatory knowledge of climate change and complex systems ideas by high school students, had a research design similar to the present study, with ABM-based learning for both experimental PF and control DI groups, yet found their PF group to have a significant advantage in learning explanatory knowledge. There is, however, a crucial difference between Jacobson et al.'s research design and that of the present study: Jacobson et al.'s instructional interventions were presented in an in-class, in-person context whereas the present study used guided learning videos in an online format. The latter condition made no allowance for face-to-face discussion whereby students would have been able to compare—interactively with the instructor—their generated solutions to the canonical solutions. Indeed, Sinha and Kapur (2019a; cf. also, Loibl & Rummel, 2014a) argue that a lack of discussion based on the students' generated solutions at the consolidation phase may reduce the effect of PF.

A related factor that might explain the diminished effect of PF for promoting explanatory knowledge in the present study is the possibility that the instructional/learning design did not include enough scaffolding (i.e., partial guidance) of concepts in the initial exploration phase, a technique that is promoted within cognitive load theory. As noted by Sweller (2010), "Factors including the complexity of the domain, learners' prior knowledge, and the quality of scaffolding can impact whether PF outperforms DI." Chen et al. (2019), for example, found that students who were presented a task without sufficient scaffolding of concepts reported a higher cognitive load and performed more poorly compared to students who were supported in these ways (see also Chen & Kalyuga, 2020).

A possible explanation may lie in the quality of the instructional materials used in the DI condition. The absence of a significant difference between the PF and DI groups in terms of declarative and explanatory knowledge may be attributed to the high-quality instructional video design used in both conditions. Unlike traditional DI, which is often characterised as passive content delivery, the DI implementation in this study included well-structured, visually engaging, and pedagogically sound videos. These materials provided clear explanations, supported retention, and reduced extraneous cognitive load—factors shown to enhance learning across diverse contexts (Andrade et al., 2015; Brame, 2016).

When both PF and DI groups received equally strong instructional scaffolding, this led to a general improvement in learning outcomes, thereby narrowing the gap typically observed between the two approaches (Ott et al., 2024). This effect may be particularly evident in the context of declarative and explanatory knowledge, where the clarity and coherence of the videos may have reduced the conceptual advantage normally associated with PF. Moreover, empirical evidence shows that when videos are particularly effective, the impact of instructional sequence—whether problem-solving precedes or follows instruction—may become less pronounced (Ott et al., 2024). This highlights that instructional quality—not merely instructional method—plays a pivotal role in learning outcomes. From a theoretical standpoint, such high-quality materials may have facilitated cognitive integration: the process through which students connect new ideas with existing knowledge structures (Ignacio, 2022; Woods et al., 2007). In interdisciplinary domains such as epidemiology, where systems thinking is essential, this form of integration is particularly valuable. Therefore, the findings of this study should be interpreted in light of the instructional scaffolding embedded in the DI condition, which likely supported deeper conceptual understanding even in the absence of exploratory problem-solving.

Thus, there are reasons to tentatively conclude that the failure to find a significant advantage of the PF condition over the DI condition in regard to the acquisition of explanatory knowledge in the present study is due to particular features of the research design. On the one hand, the strong benefits offered by ABMs in the learning process for both groups (crucially here, for the control group) may have overpowered the benefit that most previous research indicates PF has for the learning of explanatory knowledge. On the other hand, the lack of any face-to-face interaction between instructor and students in the present study (due to its online administration) and the insufficient provision of scaffolding within the PF instructional/learning design may have dampened the advantage a PF approach would normally have.

PF Versus DI for Near Within and Far Across Domain Transfer

> **RQ2:** Does the PF condition lead to superior learning outcomes in the ability to transfer knowledge of complex systems in epidemiology to new content in near within domains and in far across domains, as compared to the DI condition?

In regard to near within domain transfer (i.e., transfer of previously learned knowledge to new problems within the same domain), the results showed a significant improvement from the pretest to the post-test for the experimental condition (PF) but not for the control (DI) condition. This finding aligns with prior research indicating that giving students the opportunity to experience failure while exploring concepts prior to receiving instruction may not only promote a deeper understanding of the concepts but also strengthen learners' ability to apply those concepts in new contexts (Jacobson et al., 2015; Kapur, 2008). This result is also in line with the literature demonstrating that DI is less effective in developing conceptual understanding and knowledge transfer compared to PF (Kapur & Bielaczyc, 2012). The DI method often delivers shallow learning where students are considered receivers rather than active participants in constructing knowledge; hence, it has limited success in helping students activate their schema and solve complex problems (Jacobson et al., 2017; Keyes & Galea, 2014).

In regard to far across domain transfer as indicated by the post-test results, a significant difference was found between the two groups, with the PF group outperforming the DI group in its ability to apply knowledge of epidemics and complex systems concepts to unfamiliar contexts. This finding is in line with the research surveyed in Chapter Two that found PF promoting deep learning and knowledge transfer across domains (Jacobson et al., 2017; Kapur & Bielaczyc, 2012; Sinha & Kapur, 2021b). Jacobson et al. (2020), in particular, emphasised that integrating computational models, like ABMs, with Productive Failure pedagogy enhances students' ability to comprehend complex, dynamic systems in ways that facilitate transfer of this knowledge across different domains.

This finding is also supported by the qualitative feedback from the focus group interviews and self-reports presented in Chapter Four. As pointed out there, students in the experimental group observed how ABMs helped them see the dynamic aspects of epidemic spread in ways that traditional teaching did not. Moreover, the visual and interactive aspects of ABMs were repeatedly emphasised by students in the focus groups, who said that these features played a significant role in enhancing their understanding and ability to connect this understanding to different contexts.

The current study advances research on PF by extending its application to learning in complex systems and epidemics contexts within medical education. Compared to the predominant emphasis in tertiary education settings on PF in computer science and mathematics, the application of PF in epidemics and complex systems has received less attention (Steinhorst, 2022; Steinhorst et al., 2024). Previous PF research in higher education has mainly focused on well-structured problems based on a single discipline—often providing mixed results with respect to PF's efficacy outside mathematics. The current study evaluates ill-structured and interdisciplinary domains, that is, the multifaceted and dynamic nature of complex systems and epidemic modelling (Steinhorst et al., 2024; Zhao, 2024). This is a significant shift, as complex systems require learners to integrate reasoning about emergent phenomena and knowledge across domains, revealing distinct pedagogical and cognitive challenges compared to the more linear problem spaces typically examined in PF interventions (Zhao, 2024).

Moreover, the examination of transfer in the current study distinguishes itself from prior PF research by evaluating both far and near transfer explicitly in complex, real-world epidemics problems instead of assessing tasks involving instructional intervention (Stogniy et al., 2020). According to earlier research, PF can support transfer and conceptual understanding; however, most literature has examined transfer within discipline-specific and narrowly defined contexts (Steinhorst et al., 2024; Stogniy et al., 2020). The current study, by contrast, assesses whether PF's productive struggle improves immediate learning outcomes and enables medical students to flexibly implement their learning in broader public health challenges and novel epidemic scenarios—a potentially critical competency (Zhao, 2024; Stogniy et al., 2020).

Due to the small sample size, it is important to acknowledge the limitations in statistical power when interpreting the quantitative findings. The limited number of respondents in this study may have reduced the sensitivity to detect significant differences between the PF and DI conditions, particularly for outcome measures with small to moderate effect sizes (Gaigher et al., 2019; Hilgers et al., 2016). In educational intervention research, such limitations are common and may result in findings that appear non-significant or inconclusive, even when meaningful differences exist. With small sample sizes, conventional statistical tools could be underpowered or less reliable, which potentially impacts the generalisability and robustness of quantitative results (Hilgers et al., 2016).

Under Cohen's (1988) generic guidelines, the effect sizes in this study were interpreted as medium and small; however, these interpretations, in the context of PF research, may not capture their educational significance fully. Previous research on PF—for instance, Jacobson et al. (2020); Loibl et al. (2017); and Kapur (2010)—has associated small to medium effect sizes with meaningful gains in students' conceptual understanding and their ability to transfer learning across domains. Considering that in the current study, only three instructional sessions were involved, and a novel use of ABMs was introduced to support complex epidemiological systems learning, we can consider a small-to-moderate effect practically valuable. In educational designs, which target transferable and deep learning, this is especially true, where cognitive gains could gradually unfold while producing long-term benefits. This shows that for the PF approach in this domain, the results could be interpreted as offering promising support.

PF Versus DI and the Session-by-Session Learning Process

> **RQ3:** How does the instructional sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts affect the learning process across multiple sessions in PF vs. DI conditions?

Before evaluating the findings of RQ3 in relation to the reviewed literature, it is important to reiterate the aim of this research question (RQ3) which sought to determine whether and how the sequence of ABM-based problem-solving tasks involving complex systems and epidemiology concepts influences the learning process across multiple sessions. Specifically, it aimed to examine the impact of the learning approach (PF vs. DI) based on ABM-based problem-solving tasks involving complex systems and epidemiology concepts on the learning process and outcomes.

As explained in Chapter Three, this question was addressed by exposing students in each group to identical tasks and instructional content but presented in different sequential orders (i.e., different treatment conditions). The responses of the two groups to the tasks were coded based on the types of ideas generated, their relevance, the degree of struggle exhibited, and their awareness of surface versus structural-level features in the simulation models

Types of Ideas Generated. Drawing upon the results reported in Chapter Four, the responses of the students in the experimental group (PF) demonstrated a broader range of ideas across most sessions, often connecting seemingly unrelated concepts and producing diverse, albeit occasionally inconsistent responses. This ability to explore a wide array of perspectives suggests that PF students engaged in more creative and exploratory thinking. This aligns with the ICAP framework (Chi & Wylie, 2014), which highlights that deeper level of cognitive engagement, including the generation and testing of hypotheses, can help learners construct and refine their mental models, thereby facilitating schema-building.

Kapur's (2008) theory of iterative schema refinement within the PF framework further supports this interpretation. The PF group engaged in more diverse reasoning and generated a greater number of cross-domain analogies than the DI group. For instance, PF students compared disease spread to fire propagation, demonstrating an emerging understanding and knowledge of complex system dynamics. Although these analogies were often partially misaligned or incomplete in terms of model mechanics, they nonetheless reflect an effort to identify patterns and generalise across domains. This supports the PF hypothesis that exploration and early failure promote deeper understanding through subsequent reflection and instruction (Hartmann et al., 2020; Kapur, 2008; Loibl et al., 2017).

Moreover, even though some of the ideas were incomplete or exploratory, students from the PF group showed an ability to approach problems from multiple perspectives. The PF group, for instance, identified similarities between fire spread and disease transmission, as mentioned earlier, noting that both processes involve local interactions that lead to large-scale outcomes. This engagement in analogical reasoning reflects the early stages of schema formation and conceptual understanding. Previous research, such as Kapur and Bielaczyc (2012) and Buseyne et al. (2023), supports these findings, indicating that students in PF setting generate a wide range of ideas or solutions before receiving instruction, thereby enabling them to explore the solution space more broadly. This is in consistent with Gentner et al. (2003), who found that engaging learners in analogical comparison across domains helps abstract deep structural features which, therefore, support the generation of more diverse and transferable ideas.

On the other hand, students in the DI group provided more targeted responses within a narrower range of idea generation. While this pattern demonstrates a lack of exploratory breadth, it reveals that the DI group experienced fewer challenges as they were guided by prior instruction in solving the given problems. Their ideas, as a result, were well-articulated but did not show diversity across concepts. These findings align with Sweller (2010) and Sweller et al. (2019) regarding Cognitive Load Theory. That is, because the DI approach relies on providing instruction to students early on, it reduces the cognitive burden on their minds and working memory, helping them to present precise, well-defined ideas without much struggle. Accordingly, the DI approach, initially has greater utility than the PF approach in cases of higher element interactivity (i.e., where the lesson includes multiple concepts to be simultaneously loaded in the working memory). This claim has been forwarded as well by Loibl et al. (2017), who focused on the effectiveness of the DI approach in cases where the topic being learned has a particularly high degree of element interactivity. Furthermore, according to Chen and Kalyuga (2020), such an instruction-first approach (DI) is beneficial when procedural knowledge (e.g., solving math equations) rather than conceptual knowledge is the focus of the lesson.

Struggles and Challenges. The PF group struggled with specific interpretative and conceptual challenges. These students faced difficulties in articulating the reinfection in the SEIRS model and cyclic nature of immunity loss, and misinterpreted forest fire model's visual cues. These results are in accordance with the prediction of Kapur (2016). According to the study, in productive learning, initial failure is a crucial component as it encourages deeper cognitive processing while driving cognitive conflict. Compared to DI students, the

PF group encountered more reasoning-based and perceptual struggles, which underscores the role of exploration in identifying and resolving misconceptions.

There are some explanations for why students experience struggles under the PF approach, which include initial lack of knowledge (lack of foundational knowledge), cognitive load, and difficultly in framing failure positively (Chowrira et al., 2019; Dorland, 2023). Overall, the struggles and challenges experienced by the PF students are a natural part of the learning process, which intends to lead to more effective learning outcomes and foster deeper understanding by motivating learners to actively engage with complex problems (Chowrira et al., 2019). The findings also disclose that the DI students experience fewer struggles; however, they still face difficulties with more emergent and complex properties of the models.

Relevance of Ideas. The current findings of the present study also highlight the dimension of relevance of the ideas by both PF and DI groups. Results show that both PF and DI groups demonstrated the ability to connect learning materials to real-world situations but used different approaches to apply the models to the real-world scenarios. It was observed that PF students explored multiple perspectives but struggled with systematic reasoning and achieving deeper understanding of complex systems.

In one of the examples, PF students linked public health measures and government policies to disease spread but did not account for feedback mechanisms within the SEIRS model. This effort to engage with the complexity of the real-world applications, despite revealing gaps in understanding model-specific mechanics, aligns with the principle of productive failure, where initial struggles can lead to deeper learning outcomes (Kapur, 2008). In addition, the challenges PF students experienced in reasoning systematically about complex systems are consistent with prior findings that emphasise the importance of scaffolding and instructional support in helping learners navigate the intricate computational models (Loibl et al., 2017).

On the other hand, the DI group exhibited more accurate explanations but sometimes lacked connecting them with broader real-world implications. The DI group, in the SEIRS model, accurately described the cyclic nature of immunity and reinfection; however, they faced difficulty extending this understanding to broader epidemiological trends or policy implications. This reflects the broader distinction between conceptual and procedural knowledge. Kapur (2012) provides similar evidence that PF students generate a wider range of representations, which reflect broader exploration and more relevant ideas. This broader exploration engages multiple possible solution paths, leading to more relevant ideas.

Moreover, the PF approach helps activate the learners 'prior knowledge and identify gaps in their knowledge and understanding, which results in a more comprehensive problem-solving approach. It involves comparing multiple solutions and conceptions that are advantageous for generating relevant ideas and building conceptual knowledge (Kapur, 2008; Loibl et al., 2017; Schwartz et al., 2011).

Surface-level vs. Structural-level Understanding. The compare-contrast task of Challenge Problem 3 in the final phase of each session required all participants to compare and contrast between the two agent-based models (ABMs) they had previously worked with. The task was analysed using a coding system that classified responses as identifying surface-level features or structural-level features (Gentner et al., 2003; Jacobson et al., 2020). The results revealed a clear and consistent difference between the level of understanding of the two groups. The PF group students exhibited a deeper understanding of the underlying structural-level features than the DI group.

More specifically, across the three sessions, both PF and DI groups identified surface-level (observable) differences and similarities, such as population behaviours, system equilibrium, and spread rates of infection. The DI group, however, primarily focused on the surface-level features and did not explore the underlying mechanisms, which is a common outcome when instruction emphasises procedural or descriptive accuracy rather than conceptual understanding (Kirschner et al., 2006; Sweller et al., 2019). On the other hand, in the third session, the PF group members demonstrated a deeper understanding of the structural features and concepts, including dynamic interactions, emergent behaviour, feedback loops, and micro to macro-level transitions. This shift toward structural reasoning aligns with prior studies which indicate that productive failure approach, in conjunction with and agent-based models fosters the development of systems thinking and causal reasoning by engaging learners in exploring underlying mechanisms before receiving formal instruction (Goldstone & Wilensky, 2008; Jacobson & Wilensky, 2006; Kapur, 2014).

An evaluation of the two groups across sessions revealed that, in session 1, the PF group demonstrated a deeper understanding of how system parameters —such as resistance, latency, and immunity— influenced emergent patterns. In session 2, PF students accurately explained how based on external conditions tipping points could shift, while DI group students considered tipping points as static thresholds. Similar findings were observed in session 3 as mentioned earlier. These results are consistent with prior studies on complex systems learning, which asserts the importance of understanding the dynamic and nonlinear relationships in system behaviour (Danish et al., 2011; Goldstone & Wilensky, 2008; Jacobson et al., 2017; Jacobson & Wilensky, 2006).

Goldstone and Wilensky (2008), for instance, discuss emergent systems thinking, which requires students to develop a deeper understanding of causal structures and move beyond surface-level features. Jacobson and Wilensky (2006) argue that agent-based models facilitate learners' shift from intuitive reasoning to a more structured understanding of emergent and complex phenomena. The emphasis of the PF approach on schema-building and independent exploration appears to support this shift. On the other hand, DI's focus on procedural and terminological accuracy may limit students' comprehension to surface-level details (Kirschner et al., 2006; Sweller et al., 2011).

It is also crucial to highlight how the PF group's understanding relates to the concept of analogical comparison. Their recognition of underlying structural features aligns with prior research on analogical comparison (AC) which emphasises comparing and contrasting across to facilitate the abstraction of deeper relational principles (Gentner et al., 2003; Kurtz & Loewenstein, 2007). In the present study, the PF group's enhanced structural understanding could be attributed to the synergistic effect of productive failure and analogical comparison. Notably, students in the PF condition identified key structural factors —like feedback cycles, brand loyalty, and hospital capacity— as central mechanism influencing system tipping points and equilibrium. In contrast, although DI students recognised surface-level similarities, their responses were more descriptive and tended to causal depth.

According to Jacobson et al. (2020), structural understanding is a critical component of students' capacity for far transfer. While the DI group's focus on surface-level features may support the immediate application of knowledge within a familiar domain, it is less conducive to transferring that knowledge across domains. In contrast, the PF approach fosters a systems-level understanding that is inherently more transferable. Prior research suggests that the core design of PF approach enables students to identify deep structural relationships across problems, thereby facilitating transfer to both near and far contexts (Jacobson et al., 2015; Loewenstein et al., 1999). It appears that, in the present study, the PF students' ability to compare and contrast the two analogous ABMs can be tied to their superior performance on the far transfer task. Also, their ability to abstract structural patterns and their application across domains reflects a more developed schema for understanding complex systems.

Students in the PF group also demonstrated a deeper understanding of how to modify and control system behaviour through targeted intervention across the three sessions. This awareness of system modification shows a better grasp of control dynamics and underlying causal mechanisms. These findings support the claim that the PF approach has the potential to foster deeper insight into complex systems by promoting structural mapping and schema-building (Jacobson et al., 2020; Kapur, 2016). Compared to the DI group, PF students consistently exhibited stronger systems thinking, a greater capacity for far transfer, and more nuanced recognition of causal relationships (Loibl et al., 2017; Schwartz et al., 2011).

Correctness and depth of understanding. The correctness of the responses was analysed quantitatively through analysis of the application task associated with Challenge Problem 3. Recall that in the first two sessions, the responses of the PF and DI groups for this task were found to be insignificantly different. This means that although the students in the experimental group were observed to consistently achieve higher mean scores on the application task, the scores of the two groups were statistically equal. In session 3, however, the PF students outperformed the DI students with a significantly higher application task scores ($U = 103.5$, $p < 0.1$, $r = 0.285$). More broadly, the average scores across all three sessions were greater for the PF students than the DI students. Although the difference between these scores overall remained insignificant ($t = 1.658$, $p = 0.107$), the effect size was greater than 0.5, suggesting a medium effect that is worth considering.

Taken together, these results suggest that the PF learning approach more effectively promoted correctness of learning as exhibited on the application task. This finding aligns with previous research indicating that engaging students with challenging tasks before instruction and without immediate guidance allows them to explore and think more critically, leading to greater understanding and application of concepts compared to more conventional methods (including direct instruction) (Jacobson et al., 2017; Kapur, 2008; Steenhof et al., 2019). Thus, the PF approach, if implemented effectively, yield greater results than those achieved through DI (Kapur & Bielaczyc, 2012). Additionally, combining ABMs with these challenge problems empowers students' abilities to deal with complex systems as they try to grasp the interaction between the agents and the emergent behaviours without direct instruction (Jacobson et al., 2020; Lai et al., 2017). Therefore, the higher performance of the PF group may be attributed to the fact that they had more opportunities to refine their mental models based on their schema abstraction and the generated representation and solution methods

(RSMs).

Broader Pedagogical Implications. These findings related to RQ3 have broader pedagogical implications as well. As discussed above, the students in the PF group exhibited superior understanding of systematic principles and dynamic interactions, whereas the DI group showed limited evidence of such deeper conceptual connections. Previous research supports these findings, as the PF method has been shown to be particularly well-suited for teaching complex systems (Jacobson & Wilensky, 2006; Kapur, 2008).

The results of the present study confirm that the difference in PF and DI approaches, i.e., the difference in instructional sequencing, significantly influences learning outcomes. In the PF approach, the initial exploration phase fosters resilience and curiosity which later contribute to the achievement of deeper conceptual understanding. At the same time, providing instruction before problem-solving, as in the DI model, can support learning efficiency and procedural precision (Loibl et al., 2017). These trade-offs must be considered when considering how to best reach one's desired learning outcomes, whether that be schema-building and broad conceptual exploration or, perhaps in some cases, a narrower focus on the consolidation of domain-specific knowledge, such as the acquisition of medical terminology.

Students' Experience of Their Learning

RQ4: How do students in a PF versus DI condition experience the learning of complex systems simulated via ABMs in the context of epidemiology instruction?

This research question was evaluated via thematic analysis of qualitative data extracted from the focus groups and self-reports data. The findings provide nuanced insights into the role and effectiveness of ABMs in supporting systems thinking and conceptual understanding the comparative impact of PF and DI instructional approaches, and the challenges students face when engaging with complex models. Additionally, participants provided suggestions for improving the design and implementation of ABM-based learning and for enhancing instructional strategies through a more integrated use of PF and DI. In the following section, the findings associated with RQ4 are discussed in light of previous literature.

Engagement and Positive Learning Experiences. A positive perception of the ABMs employed in the present study was expressed by the participants from both the PF and DI groups. They reported that, for understanding complex concepts, the ABMs were visually engaging and interactive, with the ability to manipulate parameters and simulate real-life scenarios, contributing to a more practical understanding of complex systems and epidemics. These findings are consistent with previous research indicating that ABMs enhance causal and mechanistic reasoning and facilitate the visualisation of dynamic systems at both macro and micro levels (Hsiao et al., 2019; Jacobson et al., 2006; Levy & Wilensky, 2008).

The key factors that influence student engagement include clarity, structure and comprehensiveness of the instructional content. In the DI group, students appreciated the information's systematic delivery. It enabled students to understand complex concepts related to epidemiology quickly and with confidence. According to Sweller et al. (2019), structured learning enhances procedural knowledge acquisition and reduces cognitive load. The students of the DI group find it easier to follow the content due to clearly defined instructional steps and well-organised lectures. The structured format simplified the content and made it easy to understand. This aligns with previous research findings, i.e. organised content presentation improves learning efficiency (Sweller et al., 2019). On the other hand, students of the PF group struggled initially with the learning process's open-ended nature. Some respondents felt uncertain, without direct instruction, about to approach the models. They reported deeper understanding and engagement, after they adopted to exploratory format. Previous literature has also highlighted the initial struggle in the case of PF, which leads to long-term retention of knowledge and deeper conceptual understanding (Kapur, 2008; Kapur, 2016).

Visual aids, such as instructional videos and dynamic models also increase students' understanding and engagement. Students of both groups mentioned the visual models' value in improving information retention and clarifying complex epidemiological relationships. Students of the PF group, for understanding complex system behaviour and discovering patterns, used visual models as the main tool, while the DI group considered visual aids to reinforce instructors' structured explanations. Levy and Wilensky (2008) and Jacobson et al. (2006) support it, as according to them dynamic visuals help in complex problem-solving and enhance systems thinking.

Through ABMs, exploratory learning enabled the PF group to actively engage with the material. They were able to observe system outcomes and manipulate parameters, which improved problem-solving skills and encourage deeper thinking. Jacobson et al. (2017) provided similar evidence, exploration fosters long-term knowledge retention and deeper cognitive engagement. On the hand, for exploration, DI group reported limited opportunities. However, they also expressed the need for more independent experimentation. It means that direct instruction and exploration opportunities can increased understanding and engagement across both instructional models.

Challenge and Barriers to Effective Learning. ABMs allow students to engage in complex, systems-based learning; however, they also serve as barriers to learning experience due to several challenges. Participants in both groups also identified challenges related to the usability and complexity of the ABM tools. As highlighted in prior research, such barriers contribute to a steep learning curve in understanding the underlying mechanics and interpreting the outputs of ABMs (Badham et al., 2018; Helikar et al., 2015). Furthermore, in line with prior findings (Vázquez-Serrano et al., 2021), participants in the present study observed that ABMs often oversimplify real-world scenarios and may exclude critical contextual details.

Due to the absence of structured guidance and the complexity of ABMs, students from the PF group reported higher levels of cognitive load. They described feeling overwhelmed due to the complexity of system interactions and the number of parameters. With increasing complexity, the students observed to many variables to manipulate, which coincides with the observation of steep learning curve linked with ABMs (Badham et al., 2018; Helikar et al., 2015). Previous research suggests simplifying the ABMs' initial stages to address this challenge by offering foundational lessons on system behaviour and interpreting parameters (Mayer, 2019). Moreover, scaffolding strategies are recommended to reduce cognitive overload, where learners receive guidance at the beginning, followed by self-guided exploration's opportunities (Kapur, 2016).

During the learning process, the absence of structured guidance and immediate feedback is one of the significant challenges as identified by the PF group. Students felt unsure, without timely input from the instructors, whether ABM's exploration led to the right conclusion. According to Kapur (2016), PF's success is based on a well-designed instructional process, which offers sufficient help to learns encountering problems in their exploratory learning phase. Moreover, PF could be less effective for students lacking prior experience or knowledge as from their initial struggles, they are unable to extract useful insights (Loibl et al., 2017). Combining PF approach with more structured instruction could be a solution to complex problem-solving while allowing for exploration (Holmes et al., 2014).

During the learning process, both groups of students faced technical difficulties, which affected their ability to interact and engage with ABMs. These technical issues included model crashes, glitches, and slow system response times. Students from the PF group were more affected since they relied mainly on the ABMs' interactive nature. Technical difficulties such as slow model processing and system crashes decrease simulation-based learning's effectiveness (Chopra et al., 2024). Therefore, providing a contingency plan and improving system stability can reduce interruptions. Both PF and DI groups found difficulties in the question formats. Previous research argues that questions should be free from ambiguity and directly aligned with the learning objectives (Mylopoulos et al., 2018).

Comparison of Learning Methods. In the context of medical education, the comparison between PF and DI approaches as instructional strategies has been ongoing research (Portolese, 2021). Previous research suggests that PF compared to DI results in superior learning outcomes (Steenhof et al. 2019). The feedback from participants in the present study suggests that traditional instructional methods are beneficial for constructing foundational knowledge, particularly when dealing with content that imposes a heavy cognitive load or when applied to novice learners. This aligns with the argument by Loibl et al. (2017), who note that the effectiveness of PF may be reduced for new learners with limited prior knowledge, as engaging with ill-structured problems can lead to increased cognitive strain. Conversely, while the DI approach supports clarity and procedural learning, it might be less effective for fostering a deeper, more systematic understanding of complex concepts (Jacobson et al., 2017). It means that PF's effectiveness is not universal and relies on certain factors such as learners' prior knowledge, and the subject matter's complexity. It is evident that PF's success in contingent on the students' readiness to engage in problem-solving without initial support and the appropriateness of the instructional design (Kapur & Bielaczyc, 2012). Therefore, the benefits of the PF approach are not applicable to all learner groups or learning contexts.

Suggestions for Improvement. Participants from both the PF and DI groups acknowledged the potential and significance of extending ABM-based instruction to other domains, including the clinical sciences, healthcare systems, and patient behaviour—an idea also supported in prior research (Silverman et al., 2021). In line with this perspective, ABMs have previously been employed to model public health interventions (Cuevas, 2020) and to simulate health-care logistics (Narassima et al., 2020). Furthermore, participants highlighted the value of the active learning promoted by ABMs, aligning with findings by Danish et al. (2011), who revealed that complex systems models can effectively support foundational skill development in early education (also AlRuthia et al., 2019). Drawing on these insights, it is proposed that ABMs be integrated not only into the epidemiology curricula but also across other areas of medical education. Such integration could support the development of systems thinking at an early age and enhance students' conceptual understanding and preparedness for complex real-world challenges in healthcare contexts.

Integration of PF and DI approaches is discussed in this study, which can offer a balanced way to medical education. In problem-solving tasks, engaging students before giving explicit instructions enables students to identify and explore gaps in their knowledge and understanding, and can be addressed effectively with subsequent instruction. According to Steenhof et al. (2020), this method increases explanatory knowledge and application to new clinical problems. Moreover, during problem-solving phase, incorporating immediate feedback ensures that the struggles of the students result in productive learning outcomes (Mylopoulos et al., 2018).

Previous research argued that the PF approach help students to gain deeper conceptual knowledge due to the process of struggling with and exploring complex problems (Kapur, 2016). However, according to Steenhof et al. (2020), it may not be effective for learners with lower prior knowledge. Therefore, combining PF and DI methods would create a more inclusive and adaptable instructional approach.

It is also suggested to improve ABMs' usability, which is important for effective learning. Developing user-friendly interfaces can increase accessibility of these tools to both students and educators. This is also emphasised in previous studies, for example creating simulation tools that are intuitive to use, easy to navigate and provide real-time feedback can facilitate seamless learning experience and increase student engagement (Marshall et al., 2015).

The need of clear and structured guidance is discussed in this study. Previous research also supports this need; step-by-step instructions and guided tutorials can help students in understanding ABMs' functionalities (Mylopoulos et al., 2018). Moreover, reflective feedback on simulation performance can increase understanding as according to Ng et al. (2022), structured feedback when combined with real-time performance monitoring promotes effective knowledge transfer and deeper learning. It is also important to simplify question formats to achieve learning outcomes. Avoiding ambiguous terms and using straightforward language can increase students' ability to demonstrate their knowledge (Mylopoulos et al., 2018).

Lastly, thematic saturation was achieved in the qualitative analysis, as no new codes or subthemes emerged after multiple rounds of coding across the PF and DI datasets. This saturation supports the trustworthiness of the findings by indicating that the thematic categories captured the full breadth of student experiences. As detailed in the methodology chapter, the credibility of the analysis was further enhanced through reflexive journaling, peer debriefing, and partial inter-rater review, which ensured that emerging themes were grounded in the data rather than shaped by researcher bias.

Summary

This chapter has critically discussed the results of the study for each research question, in the light of previous research. The study observed no significant differences in the students' development or acquisition of declarative or explanatory knowledge between the two approaches, PF and DI. In the context of declarative knowledge, previous literature had provided similar evidence, i.e., it could be acquired without deep understanding as long as content is delivered and reinforced in a structured manner that would enable both groups to acquire it. On the other hand, a majority of studies have found that PF offers a clear advantage over the DI approach in the case of acquiring explanatory knowledge, unlike the present study. Integrating ABMs in the instructional designs of both groups is a possible explanation of there being no significant difference between the two groups in this study. That is, the inclusion of ABMs may have played a substantially positive role in the subjects' learning process, levelling the playing field between the PF and DI groups.

PF was seen to be beneficial as compared to DI in regard to near- and far-domain transfers. Previous literature has also revealed that giving students the opportunity to experience failure while exploring concepts prior to receiving instruction (i.e., the PF approach) promotes a deeper understanding of the concepts and strengthens learners' ability to apply those concepts in new contexts. The present study reinforces these findings that PF has the ability to promote deep learning and knowledge transfer across domains.

Moreover, the chapter also critically discussed the influence of difference in instructional sequence on the learning outcomes. Similar to previous studies' findings, it is observed that the exploration phase within the PF condition, fosters resilience and curiosity, which help achieve the desired learning outcomes. However, prior instructions enhance efficiency and precision, suggesting potential trade-offs between the two approaches. Lastly, the results of the thematic analysis are discussed, which supported the findings related to the effectiveness of ABMs. Considering previous research findings, the study discussed benefits of using a 'blended approach', using DI for building foundational knowledge and PF for subsequent in-depth understanding and knowledge transfer. It is further suggested to include and integrate ABMs with epidemiology and other disciplines of medicine.

Chapter Six: Conclusion

The present study investigated whether the PF and DI instructional/learning approaches, both enhanced with AMBs, positively impact learning outcomes for the instruction of epidemics and complex systems concepts. Generally speaking, the findings revealed both significant strengths and challenges associated with each instructional design as well as with the integration of ABMs and complex systems theory into an instructional design to foster deeper learning.

Regarding the efficacy of PF, compared to DI, on the types of knowledge constructed, the quantitative findings revealed no statistically significant differences between the overall performance of both groups in the post-test results when assessing declarative knowledge of epidemics, or of declarative and explanatory knowledge of complex systems in epidemiology. It was expected that the PF group could perform better than DI group in tasks relevant to explanatory knowledge, as has generally been the case in prior studies (e.g., Jacobson et al., 2017; Kapur & Bielaczyc, 2012), where PF's design can potentially better promote conceptual understanding and explanatory reasoning. Various possible factors could explain the present study's result: Most importantly, the absence of in-person interaction with an instructor in the consolidation phase may have diminished any benefit the PF approach might otherwise have offered the experimental group, and the interactive, visual nature of ABMs may have boosted the learning potential of both groups more or less equally (Goldstone & Wilensky, 2008).

The study also investigated the effects of PF versus DI on knowledge transfer, or students' ability to transfer acquired knowledge within the same subject domain or to other domains. The results revealed that the PF treatment group outperformed the DI control group on both types of transfer (a marginally significant result in the case of near within domain transfer; a significant result in the case of far across domain transfer), demonstrating that PF is more effective than a DI approach in engaging students to transfer their learned concepts of epidemics and complex systems to solve new problems whether within or outside the domain (cf. Jacobson et al., 2020; Kapur, 2008).

In addition, the study explored the session-by-session learning processes within the context of both PF and DI instructional designs to demonstrate how the sequencing of the challenge problems given in sessions may impact knowledge construction. The first two challenge problems were qualitatively analysed and revealed that PF group generated a broader range of ideas and encountered more struggles whereas the DI group showed greater focus and accuracy in their responses and with fewer struggles, as they had been exposed to the instructional content. The experience of more failure by the PF group in the exploration phase prior to receiving instruction is consistent with the literature, where it is regarded a productive step toward improved learning and knowledge transfer (Sinha & Kapur, 2021).

Theoretical and Practical Implications

As for theoretical contributions, the present study showcases the efficacy—at least insofar as facilitating knowledge transfer and the grasping of deeper structural properties of models is concerned—of utilising a Productive Failure (PF) learning design combined with Agent-Based Models (ABMs) in the teaching of epidemiology, particularly of content such as the dynamics of disease outbreaks, which involves complex systems. This instructional approach taken in this study aligns with constructivist learning theories whereby students are engaged in solving ill-structured tasks before receiving instruction. The integration of ABMs into this framework allows learners to explore the relevant interactions at both individual and systemic levels, leading to greater conceptual understanding. This work is, to this researcher's knowledge, the first to extend the literature on PF beyond its application in STEM fields and apply this pedagogical approach to the understanding of real-world public health challenges in university-level medical education.

For learning and understanding complex epidemiological systems, ABMs are considered especially useful because they operationalise key principles of complexity directly via interactive simulation. With the help of ABMs, students can manipulate individual agent behaviours and examine how macro-level patterns are generated by micro-level interactions, such as disease transmission dynamics (Tracy et al., 2018). ABMs promote understanding by exhibiting how unexpected population-level outcomes are created by simple rules (e.g., infection probabilities; Miller & Yoon, 2023), while such nonlinear relationships are often obscured by traditional static models. Adjusting parameters in real time allows students to discover stochastic effects and feedback mechanisms, which cannot be captured by deterministic models (Tracy et al., 2018). Without ABMs, learners might struggle to conceptualise how epidemics are driven by heterogeneous interactions between infected/susceptible persons or spatial network effects, potentially reducing their reasoning ability regarding threshold effects or indirect intervention impacts (Marshall & Galea, 2015). Moreover, a PF versus DI study in the absence of ABMs would have lacked a dynamic experimentation layer, limiting the capacity of students to discover system leverage points or test propositions regarding counterfactual scenarios (Miller & Yoon, 2023; Ingram, 2020).

The ABMs' exploratory nature helped enhance the effectiveness of the PF approach, as it provided an authentic way to test, before formal instruction, an alternative epidemiological mechanism (Miller & Yoon, 2023; Ingram, 2020). Future studies would benefit from a comparison between conventional diagram-based complexity instruction and ABM-enhanced curricula to determine the unique pedagogical value of ABMs. Future research might also explore hybrid methodologies, which combine system dynamics models with ABMs for teaching multiscale phenomena (Tracy et al., 2018; Hammond & Barkin, 2024). Studies could also consider using alternative visualisation tools for feedback and emergence processes, when excluding ABMs.

From a pedagogical perspective, the findings here support the recommendation that learners be engaged in problem-solving before formal instruction so as to promote their understanding of key concepts, which can be transferred to solve new, unfamiliar problems. The study further supports the idea of combining ABMs with the PF learning design in teaching epidemics and complex systems concepts, as these interactive and dynamic models enable learners to more readily observe the emergent properties that arise from these agents' interactions. The use of such computational models could substantially enhance teaching practices in health education.

Similar to Jacobson et al. (2017), who investigated how integrating complex systems together with climate change content could deepen the overall understanding of students, the present study employed a similar methodology in the field of epidemiology, aiming for robust knowledge transfer through concurrent learning of complexity and epidemics. Due to the challenges that might occur with curricular sequencing (i.e., the question of whether to teach complex systems first or specific scientific topics), the present study favours a concurrent approach where learners work on solving problems of both fields as part of the daily learning activities. Overall, this research supports a practical application of Jacobson et al.'s findings by teaching the concepts of epidemics and complexity concurrently, potentially leading to better conceptual understanding and transfer across scientific domains.

In medical education, where students must master a variety of complex diseases and biological interactions, the application of the Productive Failure (PF) approach could facilitate cross-disciplinary learning, as knowledge gained in one area can be transferred to others. As a result, this cognitive flexibility helps students make connections among distinct areas of study, apply similar problem-solving strategies across contexts, and thus better diagnose conditions by drawing on knowledge from relevant fields.

As education increasingly shifts toward online and AI-driven environments, this study's online asynchronous format takes on added significance. Absent an instructor and any immediate face-to-face discussion in such online modules, the PF approach could encourage autonomous learning facilitated by AI tools that offer instant feedback. Moreover, the present study raises the question of whether PF might be used in conjunction with the Direct Instruction (DI) method in online contexts, where PF might be used initially to promote deeper understanding, and a DI-like approach is drawn upon to provide necessary scaffolding and structure during failure. Further research is warranted to examine whether PF is adaptable to the asynchronous and remote nature of online medical education.

Productive failure (PF) highlights the effectiveness of several key competencies in medical education and other professional fields: conceptual understanding, critical thinking, and problem-solving. As learners engage with complex problems, they are required to activate their schema abstraction, synthesise, and reflect on their approaches to develop a coherent understanding of the target concepts as well as enhance their ability to transfer their knowledge to novel situations. Critical thinking is an essential component of PF, where it encourages learners to critique and adapt their strategies based on their experiences with failure, which in turn improves their problem-solving abilities. In fields like medicine, practitioners must navigate various and unpredictable clinical challenges. For instance, doctors must apply critical thinking and problem-solving skills to effectively diagnose and treat patients safely rather than rely solely on prescriptive guidelines. The use of PF in medical education could prove to be an important method for fostering such critical thinking.

Limitations and Future Research

The current study has limitations. First, it was a small-scale study and limited to a specific population of undergraduate medical students, facts which affect the generalisability of the findings to other domains within health sciences. Further research is needed that would increase the sample size and investigate whether PF can benefit other fields beyond epidemiology and across different educational levels. It is also important to acknowledge the dropout rate of participants, which reduced the sample size and lowered the statistical power of the study. This diminished power lowered the ability to detect significant differences between PF and DI conditions. According to previous research, sample attrition can affect the reliability and robustness of study outcomes, particularly in research involving educational interventions and modest effect sizes (Henneberger et al., 2023; Hoerger, 2010).

Given that this study's design required students to work individually on all the learning activities, as the study was conducted online asynchronously, future research should explore the possibility of engaging students in collaborative problem-solving using a similar learning design. As noted by Chi and Wylie (2014) (also Jacobson et al., 2017; Portolese et al., 2016), collaborative learning is enhanced through verbal interactions, for instance, discussing similarities and differences, and engaging in debates with peers. This process tends to motivate learners to engage with the study materials, which leads to more effective learning outcomes as compared to individual work.

Second, although the PF approach is regarded as a practical application of active learning principles, it runs the risk of frustrating learners. That is, PF actively immerses students in tackling complex problems before receiving instruction so that they may make mistakes and learn from their multiple failures, which subsequently fosters long-term retention and knowledge transfer (Kapur, 2016; Steenhof et al., 2019). This active engagement appears to facilitate deeper understanding and application of knowledge; however, such a process can sometimes lead to frustration during the failure phase, as evidenced by the qualitative feedback on the models' complexity (see Chapter Four). This may negatively impact students’ motivation and work progress, particularly in time-constrained environments such as medical education.

Third, and on a related note, the inclusion of ABMs into the learning design also comes with the risk of frustrating some participants, such as those in this study who reported difficulties in using the models (e.g., some found it challenging to manage and track the numerous parameters in ABMs). This may, in turn, impact the overall effectiveness of the learning experience. There were also concerns about the slow responses of the models and software bugs. All this suggests that further enhancement is needed in similar future research, such as simplifying the interface and providing more guidance to subjects. This would likely ease frustration over the use of these models and increase student engagement.

Fourth, a potential drawback of the PF teaching method, especially in medical education, is the time investment it requires. Medical domains often include a vast amount of knowledge that needs to be delivered to the students, so engaging them in problem-solving before instruction could extend the learning process, which instructors consider a disadvantage in time-constrained settings. Moreover, the application of PF may vary depending on the complexity of the task. While PF is well-suited for complex learning tasks, it could be less effective for simpler tasks, where DI may be more appropriate. Further investigation is warranted to explore the efficacy of PF in less complex or more straightforward tasks and whether prior knowledge is required in task completion.

Fifth, it is important to clarify that the findings of this study do not directly address when to use PF versus DI, as this falls out of the research scope; instead, the study sheds light on the effect of PF on the learning of complex concepts and transfer. Also, this study did not investigate the temporal dynamic of how learning evolves over a longer period of time through PF; therefore, future research needs to explore the long-term impacts and the sequencing of PF and DI in regard to the targeted concepts.

Sixth, the ABMs used in this study were not developed with the ability to collect back-end data (log file data), which could be used to assess the learning outcomes and observe students' interaction with the models. Indeed, this data collection method was initially proposed in this study, and a one-month subscription was purchased for this purpose; however, the participants unexpectedly spent closer to three months to complete all the study materials. Given the high cost of the subscription, the final completed study did not include such back-end data. Future research, however, might make good use of such data. Also, there is a need to explore the effectiveness of other types of educational technologies with PF learning design, such as 3D games and virtual environments.

Seventh, this study's current PF design was conducted online via Moodle platform, and the content materials were video recorded, so no teacher was involved to provide scaffolding and discuss the generated solutions. As noted earlier, this fact may account in part for the failure to find any significant difference between the experimental and control group’s learning of explanatory knowledge. Jacobson et al. (2017) emphasised that the teacher’s involvement in the consolidation phase of PF plays a crucial role in facilitating students’ understanding, where they can ask questions comfortably during the instruction and initiate discussions to compare their generated solutions with the canonical solutions. Therefore, future research might be considered that either conducts the study in face-to-face classrooms with a teacher or finds ways to involve synchronous discussion between the teacher and students in an online format.

Eighth, the content materials and the challenge problems solved by the students with the use of ABMs in the present study were distributed in three separate sessions. However, some of the participants complained that they needed a longer time to understand the study requirements, explore the ABMs, and observe the behaviours to answer the given questions. For future research, an alternative would be to extend the learning activity to six sessions, which would better ensure the ability of subjects to meet the study requirements successfully.

Lastly, just as Jacobson et al. (2020) compared the performance of a group of students who used a hybrid PF and analogical comparison design with two different ABMs (climate and complex systems models) to that of another group who used only the climate computer model, such a study might be usefully replicated in the field of epidemiology by comparing one group who uses only an epidemics ABM to another group who uses two models, an epidemics ABM and a complex systems ABM. The relative impact of these two instructional designs on the students' development of conceptual understanding and on their knowledge transfer could then be investigated.

Summary

In conclusion, this chapter encapsulates the main findings derived from both the qualitative and quantitative results, along with the theoretical and practical implications of the findings and the acknowledged limitations that can guide future similar studies. This study contributes to the growing understanding of how Productive Failure (PF) can be applied to professional education, particularly in the medical field, an area where it has been relatively unexplored. The existing literature commonly focuses on the utility of PF in K-12 education, but this study is one of a few that extend PF's application to the realm of adult learners in fields of knowledge involving complex systems. Specifically, this study explores how PF and Agent-Based Models (ABMs) can be used to teach complex concepts in epidemiology, such as the dynamic of epidemics diseases. The findings are generally consistent with the literature on PF and Direct Instruction (DI) methods, demonstrating that while both PF and DI are effective for teaching declarative and explanatory knowledge of epidemics and complex systems as well as enhancing the near within-domain transfer, PF highlights a significant advantage in promoting far across domain transfer.

Overall, the findings underscore the potential of PF to stimulate conceptual understanding, critical thinking, and problem-solving skills in medical education, providing a dynamic learning framework that moves beyond conventional rote learning. Presenting the benefits of PF in promoting the understanding of complex systems in epidemiology, this study suggests new avenues for future educational research in health sciences through innovative, technology-enhanced learning environments.

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. P. (2017). Agent-based modelling and simulation tools: A review of the state-of-art software. *Computer Science Review, 24*, 13–33. https://doi.org/10.1016/j.cosrev.2017.03.001

Abdel-Sater, K. A. (2011). Physiological positive feedback mechanisms. *American Journal of Biomedical Sciences, 3*(2), 145–155. https://doi.org/10.5099/aj110200145

Akpan, I. J., Shanker, M., & Offodile, O. F. (2024). Discrete-event simulation is still alive and strong: Evidence from bibliometric performance evaluation of research during COVID-19 global health pandemic. *International Transactions in Operational Research*, *31*(4), 2069–2092. https://doi.org/10.1111/itor.13418

Albert, R., & Barabási, A.L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, *74*(1), 47–97. https://doi.org/10.1103/RevModPhys.74.47

Alfieri, L., Nokes-Malach, T. J., & Schunn, C. D. (2013). Learning through case comparisons: A meta-analytic review. *Educational Psychologist, 48*(2), 87–113. https://doi.org/10.1080/00461520.2013.775712

AlRuthia, Y., Alhawas, S., Alodaibi, F., Almutairi, L., Algasem, R., Alrabiah, H. K., Sales, I., Alsobayel, H., & Ghawaa, Y. (2019). The use of active learning strategies in healthcare colleges in the Middle East. *BMC Medical Education*, *19*(1), Article 143. https://doi.org/10.1186/s12909-019-1580-4

American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action.* https://www.researchgate.net/publication/248290185_Vision_and_Change_in_Undergraduate_Biology_Education_A_Call_to_Action

Andrade, J., Huang, W. D., & Bohn, D. M. (2015). The Impact of Instructional Design on College Students' Cognitive Load and Learning Outcomes in a Large Food Science and Human Nutrition Course. *Journal of Food Science Education*, *14*(4), 127–135. https://doi.org/10.1111/1541-4329.12067

Arkin, A. P., & Schaffer, D. V. (2011). Network news: Innovations in 21st century systems biology. *Cell, 144*(6), 844–849. https://doi.org/10.1016/j.cell.2011.03.008

Ashman, G., Kalyuga, S., & Sweller, J. (2020). Problem-solving or explicit instruction: Which should go first when element interactivity is high? *Educational Psychology Review, 32*(1), 229–247. https://doi.org/10.1007/s10648-019-09500-5

Badham, J., Chattoe-Brown, E., Gilbert, N., Chalabi, Z., Kee, F., & Hunter, R. F. (2018). Developing agent-based models of complex health behaviour. *Health & Place, 54*, 170–177. https://doi.org/10.1016/j.healthplace.2018.08.022

Bai, R., Dong, W., Shi, Y., Feng, A., Xu, A., & Lyu, J. (2020). Simulation of epidemic trends for a new coronavirus under effective control measures. *New Medicine, 30*(2), 8–12. https://dx.doi.org/10.12173/j.issn.1004-5511.2020.02.03

Banack, H. R., Lesko, C. R., Whitcomb, B. C., & Kobayashi, L. C. (2021). Teaching epidemiology online (pandemic edition). *American Journal of Epidemiology, 190*(7), 1183–1189. https://doi.org/10.1093/aje/kwaa285

Barnett, S. M., & Ceci, S. J. (2002). When and Where Do We Apply What We Learn? A Taxonomy for Far Transfer. *Psychological Bulletin*, *128*(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612

Barrows, H. S. (1996). Problem-based learning in medicine and beyond: A brief overview. *New Directions for Teaching and Learning, 1996*(68), 3–12. https://doi.org/10.1002/tl.37219966804

Berglund, A. (2015). What's in a word? Concept mapping: A graphical tool to reinforce learning of epidemiological concepts. *Journal of Epidemiology and Community Health*, *69*(12), 1232–1236. https://doi.org/10.1136/jech-2014-205068

Bertaglia, G., Bondesan, A., Burini, D., Eftimie, R., Pareschi, L., & Toscani, G. (2024). New trends on the systems approach to modeling SARS-CoV-2 pandemics in a globally connected planet. *Mathematical Models and Methods in Applied Sciences, 34*(11), 1995–2054. https://doi.org/10.1142/S0218202524500301

Bjørnstad, O. N., Shea, K., Krzywinski, M., & Altman, N. (2020). The SEIRS model for infectious disease dynamics. *Nature Methods 17,* 557–558. https://doi.org/10.1038/s41592-020-0856-2

Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: A constructionist learning environment for materials science using agent-based modeling. *International Journal of Computers for Mathematical Learning, 14*, 81–119. https://doi.org/10.1007/s10758-009-9148-8

Blikstein, P., & Wilensky, U. (2010). MaterialSim: A constructionist agent-based modeling approach to engineering education. In M. J. Jacobson & P. Reimann (Eds.), *Designs for learning environments of the future: International perspectives from the learning sciences* (pp. 17–60). Springer US.

Brady, C., Holbert, N., Soylu, F., Novak, M., & Wilensky, U. (2015). Sandboxes for model-based inquiry. *Journal of Science Education and Technology, 24*(2–3), 265–286. https://doi.org/10.1007/s10956-014-9506-8

Brame, C. J. (2016). Effective Educational Videos: Principles and Guidelines for Maximizing Student Learning from Video Content. *CBE Life Sciences Education*, *15*(4), es6. https://doi.org/10.1187/cbe.16-03-0125

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking Transfer: A Simple Proposal with Multiple Implications. *Review of Research in Education*, *24*, 61–100. https://doi.org/10.3102/0091732X024001061

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Braun, V., & Clarke, V. (2022). Conceptual and design thinking for thematic analysis. *Qualitative Psychology*, *9*(1), 3–26. https://doi.org/10.1037/qup0000196

Buseyne, S., Vrijdags, A., & Raes, A. (2023). Productive Failure as a Method for Learning about Effective Collaborative Problem Solving. *International Journal of Designs for Learning*, *14*(1), 46–61. https://doi.org/10.14434/ijdl.v14i1.35221

Cao, L. (2020). *Learning genetics in game-based learning environments: Productive failure and the transfer of knowledge* [Doctoral thesis, University of Sydney]. The University of Sydney Library. https://hdl.handle.net/2123/23423

Cao, L., Lai, P. K. & Yang, H. (2024). Using productive failure to learn genetics in a game-based environment. *Instructional Science*, *52*, 309–340. https://doi.org/10.1007/s11251-023-09644-6

Cao, L., & Zhang, J. (2024). Exploring the use of productive failure to learn ecology in a virtual world: influence of self-efficacy and the use of learning strategies. *Interactive Learning Environments*, *33*(3), 2117–2135. https://doi.org/10.1080/10494820.2024.2391058

Ceberio, M., Almudí, J. M., & Franco, Á. (2016). Design and application of interactive simulations in problem-solving in university-level physics education. *Journal of Science Education and Technology, 25*(4), 590–609. https://doi.org/10.1007/s10956-016-9615-7

Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., & Chen, Y. L. (2022). Mitigating the multicollinearity problem and its machine learning approach: A review. *Mathematics, 10*(8), Article 1283. https://doi.org/10.3390/math10081283

Chan, W. K. V., Son, Y., & Macal, C. M. (2010). Agent-based simulation tutorial—Simulation of emergent behavior and differences between agent-based simulation and discrete-event simulation. In *Proceedings of the 2010 Winter Simulation Conference* (pp. 135–150). ACM Digital Library. https://doi.org/10.1109/WSC.2010.5679168

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed). Sage Publications Ltd.

Chen, D. W., Catrambone, R., Sottilare, R. A., Schwarz, J., Schwarz, J., & Sottilare, R. A. (2019). Productive failure and subgoal scaffolding in novel domains. In *Adaptive instructional systems: HCII 2019: Lecture notes in computer science* (vol. 11597, pp. 282–300). Springer International Publishing AG. https://doi.org/10.1007/978-3-030-22341-0_23

Chen, O., & Kalyuga, S. (2020). Exploring factors influencing the effectiveness of explicit instruction first and problem-solving first approaches. *European Journal of Psychology of Education, 35*(3), 607–624. https://doi.org/10.1007/s10212-019-00445-5

Chen, O., Pass, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. *Educational Psychology Review, 35*, Article 63. https://doi.org/10.1007/s10648-023-09782-w

Chi, M. T. H., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chopra, A., Rodríguez, A., Subramanian, J., Quera-Bofarull, A., Krishnamurthy, B., Prakash, B. A., & Raskar, R. (2022). Differentiable agent-based epidemiology. In *AAMAS '23: Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems* (pp. 1848–1857). ACM Digital Library. https://dl.acm.org/doi/abs/10.5555/3545946.3598851

Chopra, A., Subramanian, J., Krishnamurthy, B., & Raskar, R. (2024, May). Flame: A Framework for Learning in Agent-based ModEls. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems* (pp. 391-399).

Chou, W.-Y. S., Oh, A., & Klein, W. M. P. (2018). Addressing health-related misinformation on social media. *JAMA, 320*(23), 2417–2418. https://doi.org/10.1001/jama.2018.16865

Chowrira, S. G., Smith, K. M., Dubois, P. J., & Roll, I. (2019). DIY productive failure: boosting performance in a large undergraduate biology course. *NPJ Science of Learning*, *4*(1), 1–1. https://doi.org/10.1038/s41539-019-0040-6

Coggon, D., Rose, G., & Barker, D. J. P. (2024). What is epidemiology? In *Epidemiology for the uninitiated* (ch. 1). BMJ Publishing Group Ltd. https://thebmj-frontend.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/1-what-epidemiology

Coleman, E. B. (1998). Using explanatory knowledge during collaborative problem-solving in science. *Journal of the Learning Sciences, 7*(3-4), 387–427. https://doi.org/10.1207/s15327809jls0703&4_5

Colizza, V., Barthélemy, M., Barrat, A., & Vespignani, A. (2007). Epidemic modeling in complex realities (Réalité complexe et modèles en épidémiologie). *Comptes Rendus Biologies, 330*(4), 364–374. https://doi.org/10.1016/j.crvi.2007.02.014

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). Sage Publications.

Crooks, A., & Hailegiorgis, A. (2013). Disease modeling within refugee camps: A multi-agent systems approach. In *WSC '13: Proceedings of the 2013 Winter Simulation Conference: Making decisions in a complex world* (pp. 1697–1706). ACM Digital

Library. https://doi.org/10.1109/WSC.2013.6721551

Cuevas, E. (2020). An agent-based model to evaluate the COVID-19 transmission risks in facilities. *Computers in Biology and Medicine, 121*, Article 103827. https://doi.org/10.1016/j.compbiomed.2020.103827

Currie, C. S. M., Fowler, J. W., Kotiadis, K., Monks, T., Onggo, B. S., Robertson, D. A., & Tako, A. A. (2020). How simulation modelling can help reduce the impact of COVID-19. *Journal of Simulation, 14*(2), 83–97. https://doi.org/10.1080/17477778.2020.1751570

Danish, J., Peppler, K., & Phelps, D. (2010). BeeSign: Designing to support mediated group inquiry of complex science by early elementary students. In *IDC '10: Proceedings of the 9th International Conference on Interaction Design and Children* (pp. 182–185). https://doi.org/10.1145/1810543.1810566

Danish, J., Peppler, K., Phelps, D., & Washington, D. (2011). Life in the hive: Supporting inquiry into complexity within the zone of proximal development. *Journal of Science Education and Technology, 20*(5), 454–467. https://doi.org/10.1007/s10956-011-9313-4

de Jong, T., Lazonder, A. W., Chinn, C. A., Fischer, F., Gobert, J., Hmelo-Silver, C. E., Koedinger, K. R., Krajcik, J. S., Kyza, E. A., Linn, M. C., Pedaste, M., Scheiter, K., & Zacharia, Z. C. (2024). Beyond inquiry or direct instruction: Pressing issues for designing impactful science learning opportunities. *Educational Research Review*, *44*, Article 100623. https://doi.org/10.1016/j.edurev.2024.100623

Dickes, A. C., & Sengupta, P. (2013). Learning natural selection in 4th grade with multi-agent-based computational models. *Research in Science Education, 43*(3), 921–953. https://doi.org/10.1007/s11165-012-9293-2

Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of mechanistic reasoning and multilevel explanations of ecology in third grade using agent-based models. *Science Education, 100*(4), 734–776. https://doi.org/10.1002/sce.21217

Dietz, K. (1979). Epidemiologic interference of virus populations. *Journal of Mathematical Biology, 8*(3), 291–300. https://doi.org/10.1007/BF00276314

Diez Roux, A. V. (2002). A glossary for multilevel analysis. *Journal of Epidemiology and Community Health (1979)*, *56*(8), 588–594. https://doi.org/10.1136/jech.56.8.588

Donkin, R., Yule, H., & Fyfe, T. (2023). Online case-based learning in medical education: a scoping review. *BMC Medical Education*, *23*(1), 564–564. https://doi.org/10.1186/s12909-023-04520-w

Donnison, S., Dunn, P. K., Cole, R., Bulmer, M., Roiko, A. H., & Muller, F. (2016). Enhancing curriculum in epidemiology and biostatistics through simulation-based learning. *International Research in Education, 4*(1), 11–26. https://doi.org/10.5296/ire.v4i1.8064

Dorland, A. (2023). Failing to Learn: Design Thinking and the Development of a Failure-Positive Mindset in the University Classroom. *Collected Essays on Learning and Teaching*, *14*(1). https://doi.org/10.22329/celt.v14i1.7155

Drożdż, S., Kwapień, J., & Oświęcimka, P. (2021). Complexity in economic and social systems. *Entropy, 23*(2), Article 133. https://doi.org/10.3390/e23020133

Duch, B. J., Groh, S. E, & Allen, D. E. (Eds.). (2001). *The power of problem-based learning*. Stylus.

Earl, G. L. (2009). Using cooperative learning for a drug information assignment. *American Journal of Pharmaceutical Education, 73*(7), Article 132. https://doi.org/10.5688/aj7307132

Enworo, O. C. (2023). Application of Guba and Lincoln’s parallel criteria to assess trustworthiness of qualitative research on indigenous social protection systems. *Qualitative Research Journal*, *23*(4), 372–384. https://doi.org/10.1108/QRJ-08-2022-0116

Epstein, J. M. (2009). Modelling to contain pandemics. *Nature, 460*(7256), Article 687. https://doi.org/10.1038/460687a

Fang, Y., Nie, Y., & Penny, M. (2020). Transmission dynamics of the COVID-19 outbreak and effectiveness of government interventions: A data-driven analysis. *Journal of Medical Virology, 92*(6), 645–659. https://doi.org/10.1002/jmv.25750

Feder, T. (2007). Statistical physics is for the birds. *Physics Today, 60*(10) 28–30. https://doi.org/10.1063/1.2800090

Feltovich, P. J., Spiro, R. J., & Coulson, R. L. (1993). Learning, teaching, and testing for complex conceptual understanding. In N. Frederiksen, R. J. Mislevy, & I. I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 181–217). Lawrence Erlbaum Associates.

Gaigher, A., Burri, R., San-Jose, L. M., Roulin, A., & Fumagalli, L. (2019). Lack of statistical power as a major limitation in understanding MHC-mediated immunocompetence in wild vertebrate populations. *Molecular Ecology*, *28*(23), 5115–5132. https://doi.org/10.1111/mec.15276

Gange, S. J. (2008). Teaching epidemiologic methods. *Epidemiology, 19*(2), 353–356. https://doi.org/10.1097/EDE.0b013e318163d294

Gebbie, K., Rosenstock, L., & Hernandez, L. (2003). *Who will keep the public healthy? Educating public health professionals for the 21st century*. National Academies Press.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95*(2), 393–408. https://doi.org/10.1037/0022-0663.95.2.393

Germann, T. C., Kadau, K., Longini, I. M., & Macken, C. A. (2006). Mitigation strategies for pandemic influenza in the United States. *Proceedings of the National Academy of Sciences, 103*(15), 5935–5940. https://doi.org/10.1073/pnas.0601266103

Getz, W. M., Salter, R. M., & Sippl-Swezey, N. (2015). Using Nova to construct agent-based models for epidemiological teaching and research. In *Proceedings of the 2015 Winter Simulation Conference* (pp. 3490–3501). IEEE Xplore. https://doi.org/10.1109/WSC.2015.7408509

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. https://doi.org/0.1016/0010-0285(83)90002-6

Goldmann, E., Stark, J. H., Kapadia, F., & McQueen, M. B. (2018). Teaching epidemiology at the undergraduate level: Considerations and approaches. *American Journal of Epidemiology, 187*(6), 1143–1148. https://doi.org/10.1093/aje/kwy055

Goldstone, R. L., & Day, S. B. (2012). Introduction to "new conceptualizations of transfer of learning." *Educational Psychologist, 47*(3), 149–152. https://doi.org/10.1080/00461520.2012.695710

Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *Journal of the Learning Sciences, 17*(4), 465–516. https://doi.org/10.1080/10508400802394898

Goldwater, M. B., & Gentner, D. (2015). On the acquisition of abstract knowledge: Structural alignment and explication in learning causal system categories. *Cognition, 137*, 137–153. https://doi.org/10.1016/j.cognition.2014.12.001

Gostic, K., Gomez, A. C., Mummah, R. O., Kucharski, A. J., & Lloyd-Smith, J. O. (2020). Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *eLife*, *9*, Article e55570. https://doi.org/10.7554/eLife.55570

Grotzer, T., Kamarainen, A., Tutwiler, M., Metcalf, S., & Dede, C. (2013). Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. *BioScience, 63*(4), 288–296. https://doi.org/10.1525/bio.2013.63.4.9

Guo, X., Li, F., Yang, Z., & Dienes, Z. (2013). Bidirectional Transfer between Metaphorical Related Domains in Implicit Learning of Form-Meaning Connections. *PloS One*, *8*(7), e68100. https://doi.org/10.1371/journal.pone.0068100

Gupta, G., Kapila, R., Chopra, A., & Raskar, R. (2024). First 100 days of pandemic: An interplay of pharmaceutical, behavioral and digital interventions—A study using agent-based modeling. In *AAMAS '24: Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems* (pp. 761–770). https://dl.acm.org/doi/abs/10.5555/3635637.3662929

Gustavson, K., von Soest, T., Karevold, E., & Røysamb, E. (2012). Attrition and generalizability in longitudinal studies: findings from a 15-year population-based study and a Monte Carlo simulation study. *BMC Public Health*, *12*(1), 918–918. https://doi.org/10.1186/1471-2458-12-918

Hammond, R. A., & Barkin, S. (2024). Making evidence go further: Advancing synergy between agent-based modeling and randomized control trials. *Proceedings of the National Academy of Sciences, 121*(21), e2314993121. https://doi.org/10.1073/pnas.2314993121

Hardesty, L. (2010, February 26). *Explained: Linear and nonlinear systems.* MIT News. https://news.mit.edu/2010/explained-linear-0226

Hartmann, C., van Gog, T., & Rummel, N. (2020). Do examples of failure effectively prepare students for learning from subsequent instruction? *Applied Cognitive Psychology, 34*(4), 879–889. https://doi.org/10.1002/acp.3651

Heeren, T., & D'Agostino, R. (1987). Robustness of the two independent samples t-test when applied to ordinal scaled data. *Statistics in Medicine, 6*(1), 79–90. https://doi.org/10.1002/sim.4780060110

Helikar, T., Cutucache, C. E., Dahlquist, L. M., Herek, T. A., Larson, J. J., & Rogers, J. A. (2015). Integrating interactive computational modeling in biology curricula. *PLoS Computational Biology, 11*(3), Article e1004131. https://doi.org/10.1371/journal.pcbi.1004131

Henneberger, A. K., Rose, B. A., Feng, Y., Johnson, T., Register, B., Stapleton, L. M., Sweet, T., & Woolley, M. E. (2023). Estimating Student Attrition in School-Based Prevention Studies: Guidance from State Longitudinal Data in Maryland. *Prevention science*, *24*(5), 1035–1045. https://doi.org/10.1007/s11121-023-01533-1

Hilgers, R.-D., Roes, K., & Stallard, N. (2016). Directions for new developments on statistical design and analysis of small population group trials. *Orphanet Journal of Rare Diseases*, *11*, 1-10. https://doi.org/10.1186/s13023-016-0464-5

Hirsch, G., Levine, R., & Weaver, M. (2023). Mapping complex public health problems with causal loop diagrams. *International Journal of Epidemiology, 53*(4), Article dyae091. https://doi.org/10.1093/ije/dyae091

Hoerger M. (2010). Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. *Cyberpsychology, behavior and social networking*, *13*(6), 697–700. https://doi.org/10.1089/cyber.2009.0445

Hoertel, N., Blachier, M., Blanco, C., Olfson, M., Massetti, M., Sánchez Rico, M., Limosin, F., & Leleu, H. (2020). A stochastic agent-based model of the SARS-CoV-2 epidemic in France. *Nature Medicine 26*, 1417–1421. https://doi.org/10.1038/s41591-020-1001-6

Holland, J. H. (1992). Genetic Algorithms. *Scientific American*, *267*(1), 66–73. https://doi.org/10.1038/scientificamerican0792-66

Holmes, N. G., Day, J., Park, A. H. K., Bonn, D. A., & Roll, I. (2014). Making the failure more productive: scaffolding the invention process to improve inquiry behaviors and outcomes in invention activities. *Instructional Science*, *42*(4), 523–538. https://doi.org/10.1007/s11251-013-9300-7

Holbert, N., & Wilensky, U. (2014). Constructible authentic representations: Designing video games that enable players to utilize knowledge developed in-game to reason about science. *Technology, Knowledge and Learning, 19*, 53–79. https://doi.org/10.1007/s10758-014-9214-8

Hossain, M. M. (2022). *Teaching epidemiology: An overview of strategies and considerations* [Preprint]. SSRN. https://doi.org/10.2139/ssrn.4141332

HosseiniKhezri, A., Shang, C., Moss, A. C., Song, J., & Chen, A. (2025). Learning in Physical Education: A Transfer of Learning Theory Perspective. *Journal of Teaching in Physical Education*, *1*(aop), 1-8. https://doi.org/10.1123/jtpe.2024-0251

Hsiao, L., Lee, I., & Klopfer, E. (2019). Making sense of models: How teachers use agent-based modeling to advance mechanistic reasoning. *British Journal of Educational Technology, 50*(5), 2203–2216. http://dx.doi.org/10.1111/bjet.12844

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis* (2nd ed.). John Wiley & Sons.

Hunter, E., Namee, B., & Kelleher, J. (2017). A taxonomy for agent-based models in human infectious disease epidemiology. *Journal of Artificial Societies and Social Simulation, 20*(3), Article 2. https://doi.org/10.18564/jasss.3414

Hussein, A. (2009). The use of triangulation in social sciences research: Can qualitative and quantitative methods be combined? *Journal of Comparative Social Work, 4*(1), 106–117. https://doi.org/10.31265/jcsw.v4i1.48

Ignacio, J., Chen, H.-C., & Roy, T. (2022). Advantages and challenges of fostering cognitive integration through virtual collaborative learning: a qualitative study. *BMC Nursing*, *21*(1), Article 251. https://doi.org/10.1186/s12912-022-01026-6

Ingram, F. J. (2020). *An agent-based classroom lessons model and simulation* (Unpublished doctoral dissertation). Lancaster University. Retrieved from https://eprints.lancs.ac.uk/id/eprint/159590/

Ivorra, B., Ferrández, M. R., Vela-Pérez, M., & Ramos, A. M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation, 88*, Article 105303. https://doi.org/10.1016/j.cnsns.2020.105303

Jacobson, M. J. (2001). Problem solving, cognition, and complex systems: Differences between experts and novices. *Complexity, 6*(3), 41–49. https://doi.org/10.1002/cplx.1027

Jacobson, M. J., Goldwater, M., Markauskaite, L., Lai, P. K., Kapur, M., Roberts, G., & Hilton, C. (2020). Schema abstraction with productive failure and analogical comparison: Learning designs for far across domain transfer. *Learning and Instruction*, *65*, Article 101222. https://doi.org/10.1016/j.learninstruc.2019.101222

Jacobson, M. J., & Markauskaite, L. (2015, April 16–20). *Understanding complex systems and climate change: Learning designs with agent-based models, productive failure, and analogical encoding* [Conference presentation]. Annual Meeting of the American Educational Research Association, Chicago, IL. https://www.aera.net/Publications/Online-Paper-Repository/AERA-Online-Paper-Repository

Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 52*, 1–14. https://doi.org/10.1016/j.learninstruc.2017.03.007

Jacobson, M. J., Taylor, C. E., & Richards, D. (2016). Computational scientific inquiry with virtual worlds and agent-based models: New ways of doing science to learn science. *Interactive Learning Environments, 24*(8), 2080–2108. https://doi.org/10.1080/10494820.2015.1079723

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences, 15*(1), 11–34. https://doi.org/10.1207/s15327809jls1501_4

James, E. L., Graham, M. L., Snow, P. C., & Ward, B. M. (2006). Teaching research and epidemiology to undergraduate students in the health sciences. *Australian and New Zealand Journal of Public Health*, *30*(6), 575–578. https://doi.org/10.1111/j.1467-842X.2006.tb00790.x

JASP Team. (2025). *JASP* (Version 0.19.3) [Computer software]. https://jasp-stats.org/download/

Jeffreys, H. (1961). *Theory of Probability* (3rd ed.). Oxford, UK: Oxford University Press.

Kalet, A., Ellaway, R. H., Song, H. S., Nick, M., Sarpel, U., Hopkins, M. A., Hill, J., Plass, J. L., & Pusic, M. V. (2013). Factors influencing medical student attrition and their implications in a large multi-center randomized education trial. *Advances in Health Sciences Education: Theory and Practice*, *18*(3), 439–450. https://doi.org/10.1007/s10459-012-9382-z

Kalyuga, S. (2015). *Instructional guidance: A cognitive load perspective.* IAP Information Age Publishing.

Kapur, M. (2008). Productive failure. *Cognition and Instruction, 26*(3), 379–424. https://doi.org/10.1080/07370000802212669

Kapur, M. (2010). Productive failure in mathematical problem solving. *Instructional Science,* *38*(6), 523–550. https://doi.org/10.1007/s11251-009-9093-x

Kapur, M. (2014). Productive failure in learning math. *Cognitive Science, 38*(5), 1008–1022. https://doi.org/10.1111/cogs.12107

Kapur, M. (2016). Examining productive failure, productive success, unproductive failure, and unproductive success in learning. *Educational Psychologist, 51*(2), 289–299. https://doi.org/10.1080/00461520.2016.1155457

Kapur, M. (2024). *Productive failure: Unlocking deeper learning through the science of failing* (1st ed.). John Wiley & Sons.

Kapur, M., & Bielaczyc, K. (2012). Designing for productive failure. *Journal of the Learning Sciences*, *21*(1), 45–83. https://doi.org/10.1080/10508406.2011.591717

Karunarathna, I., Gunasena, P., Hapuarachchi, T., & Gunathilake, S. (2024). *Comprehensive data collection: Methods, challenges, and the importance of accuracy.* ResearchGate. http://dx.doi.org/10.13140/RG.2.2.13134.47689

Keselman, H. J., Gaines, P. A., & Clinch, J. J. (1979). Tests for homogeneity of variance. *Communications in Statistics-Simulation and Computation, 8*(2), 113–129. https://doi.org/10.1080/03610917908812108

Kass, R. E., & Raftery, A. E. (1995). Bayes Factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572

Keyes, K. M., & Galea, S. (2014). Current practices in teaching introductory epidemiology: How we got here, where to go. *American Journal of Epidemiology, 180*(7), 661–668. https://doi.org/10.1093/aje/kwu219

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE guide no. 131. *Medical Teacher*, *42*(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Kim, B., Pathak, S. A., Jacobson, M. J., Zhang, B., & Gobert, J. D. (2015). Cycles of exploration, reflection, and consolidation in model-based learning of genetics. *Journal* *of Science Education and Technology, 24*(6), 789–802. https://doi.org/10.1007/s10956-015-9564-6

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kitano, H. (2002). Computational systems biology. *Nature, 420*, 206–210. https://doi.org/10.1038/nature01254

Kolchraiber, F. C., de Olivera Freitas, M. A., de Santana, C. L. A., Hino, P., de Souza, K. M. J., & Gamba, M. A. (2019). Pedagogical strategy for teaching and learning epidemiology in nursing undergraduate school. *Revista Brasileira de Enfermagem*, *72*(2). https://doi.org/10.1590/0034-7167-2018-0077

Koppal, M., & Caldwell, A. (2004). Meeting the challenge of science literacy: Project 2061 efforts to improve science education. *Cell Biology Education, 3*(1), 28–30. https://doi.org/10.1187/cbe.03-10-0016

Kouzy, R., Abi Jaoude, J., Kraitem, A., El Alam, M. B., Karam, B., Adib, E., Zarka, J., Traboulsi, C., Akl, E. W., & Baddour, K. (2020). Coronavirus goes viral: Quantifying the COVID-19 misinformation epidemic on Twitter. *Cureus*, *12*(3), Article e7255. https://doi.org/10.7759/cureus.7255

Krieger, N. (2012). Who and what is a "population"? Historical debates, current controversies, and implications for understanding "population health" and rectifying health inequities. *Milbank Quarterly, 90*(4), 634–681. https://doi.org/10.1111/j.1468-0009.2012.00678.x

Kurtz, K. J., & Loewenstein, J. (2007). Converging on a new role for analogy in problem solving and retrieval: When two problems are better than one. *Memory & Cognition,* *35*(2), 334–341. https://doi.org/10.3758/BF03193454

Lai, P. K., Portolese, A., & Jacobson, M. J. (2017). Does sequence matter? Productive failure and designing online authentic learning for process engineering. *British Journal of Educational Technology*, *48*(6), 1217–1227. https://doi.org/10.1111/bjet.12492

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science, 1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Last, J. M. (2001). *Dictionary of epidemiology* (4th ed.) Oxford University Press.

Laszlo, A., & Krippner, S. (1998). Systems theories: Their origins, foundations, and development. In J. S. Jordan (Ed.), *Advances in psychology: Systems theories and a priori aspects of perception* (vol. 126, pp. 47–74). Elsevier. http://dx.doi.org/10.1016/S0166-4115(98)80017-4

Lee, D., & Lee, Y. (2024). Productive Failure-based Programming Course to Develop Computational Thinking and Creative Problem-Solving Skills in a Korean Elementary School. *Informatics in Education*, *23*(2), 385–408. https://doi.org/10.15388/infedu.2024.14

Lemke, J. L., & Sabelli, N. H. (2008). Complex systems and educational change: Towards a new research agenda. *Educational Philosophy and Theory, 40*(1), 118–129. https://doi.org/10.1111/j.1469-5812.2007.00401.x

Levin, S. A. (1998). Ecosystems and the Biosphere as Complex Adaptive Systems. *Ecosystems (New York)*, *1*(5), 431–436. https://doi.org/10.1007/s100219900037

Levy, S. T, & Wilensky, U. (2009). Students' learning with the connected chemistry (CC1) curriculum: Navigating the complexities of the particulate world. *Journal of Science Education and Technology, 18*(3), 243–254. https://doi.org/10.1007/s10956-009-9145-7

Levy, S. T., & Wilensky, U. (2008). Inventing a "mid level" to make ends meet: Reasoning between the levels of complexity. *Cognition and Instruction, 26*(1), 1–47. https://doi.org/10.1080/07370000701798479

Li, M. Y. (2018). *An introduction to mathematical modeling of infectious diseases*. Springer International Publishing.

Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry.* Sage Publications.

Loewenstein, J., Thompson, L., & Gentner, D. (1999). Analogical encoding facilitates knowledge transfer in negotiation. *Psychonomic Bulletin & Review, 6*(4), 586–597. https://doi.org/10.3758/BF03212967

Loibl, K., & Rummel, N. (2014a). The impact of guidance during problem-solving prior to instruction on students' inventions and learning outcomes. *Instructional Science, 42*(3), 305–326. https://doi.org/10.1007/sl 1251-013-9282-5

Loibl, K., & Rummel, N. (2014b). Knowing what you don't know makes failure productive. *Learning and Instruction, 34*, 74–85. https://doi.org/10.1016/j.learninstruc.2014.08.004

Loibl, K., Roll, I., & Rummel, N. (2017). Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review, 29*(4), 693–715. https://doi.org/10.1007/s10648-016-9379-x

Markauskaite, L., & Jacobson, M. J. (2016). Tracking and assessing students' learning strategies in model-based learning environments. In P. Reimann, S. Bull, M. Kickmeier-Rust, R. Vatrapu, & B. Wasson (Eds.), *Measuring and visualizing learning in the information-rich classroom* (ch. 10). Routledge. https://doi.org/10.4324/9781315777979

Marshall, D. A., Burgos-Liz, L., IJzerman, M. J., Osgood, N. D., Padula, W. V., Higashi, M. K., Wong, P. K., Pasupathy, K. S., & Crown, W. (2015). Applying Dynamic

Simulation Modeling Methods in Health Care Delivery Research—The SIMULATE Checklist: Report of the ISPOR Simulation Modeling Emerging Good Practices Task Force. *Value in Health*, *18*(1), 5–16. https://doi.org/10.1016/j.jval.2014.12.001

Marshall, B. D. L., & Galea, S. (2015). Formalizing the Role of Agent-Based Modeling in Causal Inference and Epidemiology. *American Journal of Epidemiology*, *181*(2), 92–99. https://doi.org/10.1093/aje/kwu274

Mayer, R. E. (2019). How multimedia can improve learning and instruction. In J. Dunlosky & K. A. Rawson (Eds.), *The Cambridge handbook of cognition and education* (pp. 460–479). Cambridge University Press. https://doi.org/10.1017/9781108235631.019

Mazziotti, C., Loibl, K., & Rummel, N. (2015). *Collaborative or individual learning within productive failure: Does the social form of learning make a difference?* In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Exploring the material conditions of learning: The Computer Supported Collaborative Learning (CSCL) Conference 2015, Volume 2* (pp. 570–575). Gothenburg, Sweden: International Society of the Learning Sciences. https://repository.isls.org/handle/1/468

Mazziotti, C., Rummel, N., Deiglmayr, A., & Loibl, K. (2019). Probing boundary conditions of productive failure and analyzing the role of young students' collaboration. *NPJ Science of Learning*, *4*, Article 2. https://doi.org/10.1038/s41539-019-0041-5

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282. https://doi.org/10.11613/bm.2012.031

Miksch, F., Jahn, B., Espinosa, K. J., Chhatwal, J., Siebert, U., & Popper, N. (2019). Why should we apply ABM for decision analysis for infectious diseases?—An example for dengue interventions. *PloS One, 14*(8), Article e0221564. https://doi.org/10.1371/journal.pone.0221564

Mil, M. H. W., Postma, P. A., Boerwinkel, D. J., Klaassen, K., & Waarlo, A. J. (2016). Molecular mechanistic reasoning: Toward bridging the gap between the molecular and cellular levels in life science education. *Science Education, 100*(3), 517–585. https://doi.org/10.1002/sce.21215

Miller, K. M., & Yoon, S. A. (2023). Teaching complexity in biology through agent-based simulations: the relationship between students' knowledge of complex systems and metamodeling knowledge. *Frontiers in Education*, *8*, 1198307. https://doi.org/10.3389/feduc.2023.1198307

Montefusco, A., & Angeli, F. (2024). Turning complexity into a Delight to the Mind: An integrative framework for teaching and learning complex reasoning. *Management Learning*. https://doi.org/10.1177/13505076241258932

Moshe, I., Terhorst, Y., Paganini, S., Schlicker, S., Pulkki-Råback, L., Baumeister, H., Sander, L. B., & Ebert, D. D. (2022). Predictors of Dropout in a Digital Intervention for the Prevention and Treatment of Depression in Patients with Chronic Back Pain: Secondary Analysis of Two Randomized Controlled Trials. *Journal of Medical Internet Research*, *24*(8), e38261–e38261. https://doi.org/10.2196/38261

Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R., Massari, M., Salmaso, S., Tomba, G. S., Wallinga, J., Heijne, J., Sadkowska-Todys, M., Rosinska, M., & Edmunds, W. J. (2008). Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine*, *5*(3), Article e74. https://doi.org/10.1371/journal.pmed.0050074

Mudd, A. L., Bal, M., Verra, S. E., Poelman, M. P., & Kamphuis, C. (2024). Analysis of how a complex systems perspective is applied in studies on socioeconomic inequalities in health and health behaviour—a call for reporting guidelines. *Health Research Policy* *and Systems*, *22*(1), 1-9. https://doi.org/10.1186/s12961-024-01248-x

Muellner, U., Fournié, G., Muellner, P., Ahlstrom, C., & Pfeiffer, D. U. (2018). Epidemix—An interactive multi-model application for teaching and visualizing infectious disease transmission. *Epidemics, 23*, 49–54. https://doi.org/10.1016/j.epidem. 2017.12.003

Murray, J. D. (2007). *Mathematical biology: I. An introduction* (3rd ed., Vol. 17). Springer.

Mylopoulos, M., Steenhof, N., Kaushal, A., & Woods, N. N. (2018). Twelve tips for designing curricula that support the development of adaptive expertise. *Medical Teacher*, *40*(8), 850–854. https://doi.org/10.1080/0142159X.2018.1484082

Naimi, A. I. (2019). Commentary: Integrating complex systems thinking into epidemiologic research. *Epidemiology, 27*(6), 843–847. https://doi.org/10.1097/ede.0000000000000538

Narrassima, M. S., Anbuudayasankar, S. P., Jammy, G. R., Pant, R., Choudhury, L., Ramakrishnan, A., & John, D. (2021). An agent based model for assessing transmission dynamics and health systems burden for COVID-19. *Indonesian Journal of Electrical Engineering and Computer Science 24*(3), 1735–1743. http://doi.org/10.11591/ijeecs.v24.i3.pp1735-1743

National Research Council. (2013). *Next generation science standards: For states, by states.* The National Academies Press. https://doi.org/10.17226/18290

Nelson, A. L., Bradley, L., & MacDonald, P. D. M. (2018). Designing an interactive field epidemiology case study training for public health practitioners. *Frontiers in Public Health*, *6*, 275. https://doi.org/10.3389/fpubh.2018.00275

Neuman, W. L. (2014). *Social research methods: qualitative and quantitative approaches*. Pearson.

Nicholas, C. G., & Christophe, F. (2008). Mathematical models of infectious disease transmission. *Nature Reviews Microbiology, 6*(6), 477–487.

https://doi.org/10.1038/nrmicro1845

Niu, X., Zhang, J., Xu, K. M., & Wang, X. (2021). The Impact of Productive Failure on Learning Performance and Cognitive Load: Using Hypervideo to Facilitate Online Interactions. *2021 International Conference on Advanced Learning Technologies (ICALT)*, 30–32. https://doi.org/10.1109/ICALT52272.2021.00016

Ng, S. L., Forsey, J., Boyd, V. A., Friesen, F., Langlois, S., Ladonna, K., Mylopoulos, M., & Steenhof, N. (2022). Combining adaptive expertise and (critically) reflective practice to support the development of knowledge, skill, and society. *Advances in Health Sciences Education: Theory and Practice*, *27*(5), 1265–1281. https://doi.org/10.1007/s10459-022-10178-8

O'Sullivan, D., Evans, T., Manson, S., Metcalf, S., Ligmann-Zielinska, A., & Bone, C. (2016). Strategic directions for agent-based modeling: Avoiding the YAAWN syndrome. *Journal of Land Use Science, 11*(2), 177–187. https://doi.org/10.1080/1747423X.2015.1030463

Ofstad, W., & Brunner, L. J. (2013). Team-based learning in pharmacy education. *American Journal of Pharmaceutical Education, 77*(4), Article 70. https://doi.org/10.5688/ajpe77470

Okunlola, J. O. (2023). Learning transfer in the workplace: an insight into the missing link in the education and training of employees. *Studies in Learning and Teaching*, *4*(2), 349-354. https://doi.org/10.46627/silet.v4i2.241

Olsen, J., Greene, N., Saracci, R., & Trichopoulos, D. (Eds.). (2015). *Teaching epidemiology: A guide for teachers in epidemiology, public health and clinical medicine* (4th ed.). Oxford University Press.

Olshan, A. F., Diez Roux, A. V., Hatch, M., & Klebanoff, M. A. (2019). Epidemiology: Back to the Future. *American Journal of Epidemiology*, *188*(5), 814–817.

https://doi.org/10.1093/aje/kwz045

Ormand, C. (2010, February 17). *What constitutes a complex system?* National Association of Geoscience Teachers. https://serc.carleton.edu/NAGTWorkshops/complexsystems/introduction.html

Ott, T., Demare, T., Möhrke, J., Silber, S., Schwab, J., Reuter, L., Westhphal, R., Schmidtmann, I., Dietz, S.-O., Pirlich, N., Ziebart, A., & Engelhard, K. (2024). Does an instructional video as a stand-alone tool promote the acquisition of practical clinical skills? A randomised simulation research trial of skills acquisition and short-term retention. *BMC Medical Education*, *24*(1), 714. https://doi.org/10.1186/s12909-024-05714-6

Page, S. E., & Zelner, J. (2020). Population health as a complex adaptive system of systems. In Y. Apostolopoulus, M. K. Lemke., & K. H. Lich (Eds.), *Complex systems and population health: A primer* (pp. 33–44). Oxford University Press. https://doi.org/10.1093/oso/9780190880743.003.0003

Palominos, E., Levett-Jones, T., Power, T., & Martinez-Maldonado, R. (2022). 'We learn from our mistakes': Nursing students' perceptions of a productive failure simulation. *Collegian Journal of the Royal College of Nursing Australia*, *29*(5), 708–712. https://doi.org/10.1016/j.colegn.2022.02.006

Panthakkalakath, Z. E., Neeraj, N., & Mathew, J. (2023). *A framework for modeling, analyzing, and decision-making in disease spread dynamics and medicine/vaccine distribution.* arXiv. https://doi.org/10.48550/arXiv.2311.09984

Parker, A. M., Vardavas, R., Marcum, C. S., & Gidengil, C. A. (2013). Conscious consideration of herd immunity in influenza vaccination decisions. *American Journal of Preventive Medicine, 45*(1), 118–121. https://doi.org/10.1016/j.amepre.2013.02.016

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students: Findings and insights from twenty years of research*. Jossey-Bass.

Pituch, K. A., & Stevens, J. P. (2015). Assumptions in MANOVA. In *Applied multivariate statistics for the social sciences: Analyses with SAS and IBM's SPSS* (6th ed., pp. 219–264). Routledge.

Porta, M. (Ed.). (2008). *A dictionary of epidemiology* (5th ed.). Oxford University Press.

Portolese, A. E. (2021). *Productive failure in medical education: Addressing issues in problem-based learning by improving consolidation* [Doctoral thesis, University of Sydney]. The University of Sydney Library. https://hdl.handle.net/2123/26766

Portolese, A., Markauskaite, L., Lai, P. K., & Jacobson, M. J. (2016). *Analyzing patterns of emerging understanding and misunderstanding in collaborative science learning: A method for unpacking critical turning points.* In C. K. Looi, U. Cress, J. Polman, & P. Reimann (Eds.), *Transforming learning, empowering learners: The International Conference of the Learning Sciences (ICLS) 2016* (vol. 1, pp. 410–417). International Society of the Learning Science. https://repository.isls.org/bitstream/1/143/1/54.pdf

Pourshanazari, A. A., Roohbakhsh, A., Khazaei, M. R., & Tajadini, H. (2013). Comparing the long-term retention of a physiology course for medical students with the traditional and problem-based learning. *Advances in Health Sciences Education*, *18,* 91–97. https://doi.org/10.1007/s10459-012-9357-0

Rahmani, A. M., Groot, W., & Rahmani, H. (2024). Dropout in online higher education: a systematic literature review. *International Journal of Educational Technology in Higher Education*, *21*(1), 19–24. https://doi.org/10.1186/s41239-024-00450-9

Reperant, L. A., & Osterhaus, A. D. M. E. (2017). AIDS, avian flu, SARS, MERS, Ebola, Zika… what next? *Vaccine*, *35*(Pt. A), 4470–4474. https://doi.org/10.1016/j.vaccine.2017.04.082

Roberts, H., Seymour, B., Fish, S. A., Robinson, E., & Zuckerman, E. (2017). Digital health communication and global public influence: A study of the Ebola epidemic. *Journal* *of Health Communication*, *22*(sup1), 51–58. https://doi.org/10.1080/10810730.2016.1209598

Robertson, C., Safta, C., Collier, N., Ozik, J., & Ray, J. (2024). *Bayesian calibration of stochastic agent-based model via random forest.* arXiv. https://doi.org/10.48550/arXiv.2406.19524

Roll, I., Butler, D., Yee, N., Welsh, A., Perez, S., Briseno, A., Perkins, K., & Bonn, D. (2018). Understanding the impact of guiding inquiry: The relationship between directive support, student attributes, and transfer of knowledge, attitudes, and behaviours in inquiry learning. *Instructional Science, 46*(1), 77–104. https://doi.org/10.1007/s11251-017-9437-x

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science*, *7*(10), 201199-. https://doi.org/10.1098/rsos.201199

Saeedian, M., Azimi-Tafreshi, N., Jafari, G. R., & Kertész, J. (2017). Epidemic spreading on evolving signed networks. *Physical Review E, 95*(2), 022314. https://doi.org/10.1103/PhysRevE.95.022314

Samon, S., & Levy, S. T. (2017). Micro-macro compatibility: When does a complex systems approach strongly benefit science learning? *Science Education, 101*(6), 985–1014. https://doi.org/10.1002/sce.21301

Schwartz, D. L., Chase, C. C., Oppezzo, M. A., & Chin, D. B. (2011). Practicing versus inventing with contrasting cases: The effects of telling first on learning and transfer. *Journal of Educational Psychology, 103*(4), 759–775. https://doi.org/10.1037/a0025140

Sedgwick, P. (2012). Multiple significance tests: The Bonferroni correction. *British Medical*

Journal, *344,* Article e509. https://doi.org/10.1136/bmj.e509

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies, 18*(2), 351–380. https://doi.org/10.1007/s10639-012-9240-x

Sengupta, P., & Wilensky, U. (2009). Learning electricity with NIELS: Thinking with electrons and thinking in levels. *International Journal of Computers for Mathematical Learning, 14*(1), 21–50. https://doi.org/10.1007/s10758-009-9144-z

Shatnawi, M., Lazarova-Molnar, S., & Zaki, N. (2013, March 17-19). *Modeling and simulation of epidemic spread: Recent advances* [Conference presentation]. 9th International Conference on Innovations in Information Technology (IIT), Abu Dhabi, United Arab Emirates. https://doi.org/10.1109/Innovations.2013.6544404

Silva, P. C. L., Batista, P. V. C., Lima, H. S., Alves, M. A., Guimarães, F. G., & Silva, R. C. P. (2020). COVID-ABS: An agent-based model of COVID-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals, 139,* Article 110088. https://doi.org/10.1016/j.chaos.2020.110088

Silverman, E., Gostoli, U., Picascia, S., Almagor, J., McCann, M., Shaw, R., & Angione, C. (2021). Situating agent-based modelling in population health research. *Emerging Themes in Epidemiology, 18*(1), Article 10. https://doi.org/10.1186/s12982-021-00102-7

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill*. Harvard University Press.

Sinha, T., & Kapur, M. (2019). *When productive failure fails* [Conference presentation]. 41st Annual Conference of the Cognitive Science Society, Montreal, Quebec, Canada. https://www.researchgate.net/publication/333005127_When_Productive_Failure_Fail s

Sinha, T., & Kapur, M. (2021). When problem solving followed by instruction works: Evidence for productive failure. *Review of Educational Research*, *91*(5), 761–798. https://doi.org/10.3102/00346543211019105

Soros, G. (2008). *The New Paradigm for Financial Markets: The Credit Crisis of 2008 and What It Means* (1st ed.). PublicAffairs.

Steenhof, N., Woods, N. N., & Mylopoulos, M. (2020). Exploring why we learn from productive failure: insights from the cognitive and learning sciences. *Advances in Health Sciences Education: Theory and Practice*, *25*(5), 1099–1106. https://doi.org/10.1007/s10459-020-10013-y

Steenhof, N., Woods, N. N., Van Gerven, P. W., & Mylopoulos, M. (2019). Productive failure as an instructional approach to promote future learning. *Advances in Health Sciences Education*, *24*(4), 739–749. https://doi.org/10.1007/s10459-019-09895-4

Stefano, M., Hao, H., Colizza Vittoria, J. J., Duygu, B., Bruno, G., Ajelli, M., & Alessandro, V. (2010). Comparing large-scale computational approaches to epidemic modeling: Agent-based versus structured metapopulation models. *BMC Infectious Diseases, 10*(1), Article 190. https://doi.org/10.1186/1471-2334-10-190

Steinhorst, P. (2022). Investigating Productive Failure in Computer Science. *In Proceedings of the 2022 ACM Conference on International Computing Education Research-Volume 2* (pp. 19-20). https://doi.org/10.1145/3501709.3544300

Steinhorst, P., Duhme, C., Jiang, X., & Vahrenhold, J. (2024). Recognizing patterns in productive failure. In *SIGCSE 2024: Proceedings of the 55th ACM Technical Symposium on Computer Science Education* (vol. 1, pp. 1293–1299). ACM Digital Library. https://doi.org/10.1145/3626252.3630915

Steinhorst, P., Petersen, A., Simion, B., Vahrenhold, J., Fisler, K., Franklin, D., Hamilton, M., & Denny, P. (2023). Exploring Barriers in Productive Failure. *Proceedings of the 2023 ACM Conference on International Computing Education Research - Volume 1*, 284–297. https://doi.org/10.1145/3568813.3600111

Steinhorst, V., Serova, K., & Rummel, N. (2020). When failure fails to be productive: Probing the effectiveness of productive failure for learning beyond STEM domains. *Instructional Science, 48*(6), 651–697. https://doi.org/10.1007/s11251-020-09525-2

Stockburger, D., & Frey, B. B. (2018). Multivariate analysis of variance. In *The Sage encyclopedia of educational research, measurement, and evaluation* (vol. 4, pp. 1119–1125). Sage Publications. https://doi.org/10.4135/9781506326139.n456

Stogniy, O., Halmo, S. M., Reinhart, P., Alele, V., Snuggs, G., Fiorella, L., & Lemons, P. P. (2020). Comparing the Impacts of Evidence-Based Pedagogies in Undergraduate Biochemistry: Lessons from Productive Failure, Worked Examples Plus Practice, and Guided Inquiry. *The FASEB Journal*, *34*(S1), 1-1. https://doi.org/10.1096/fasebj.2020.34.s1.04148

Strobel, J., & Barneveld, A.V. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-based Learning, 3,* 44–58. https://doi.org/10.7771/1541-5015.1046

Suriyaarachchi, H., Denny, P., & Nanayakkara, S. (2025). Investigating the Use of Productive Failure as a Design Paradigm for Learning Introductory Python Programming. *In Proceedings of the 56th ACM Technical Symposium on Computer Science Education V. 1* (pp. 1085-1091). https://doi.org/10.1145/3641554.3701911

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Van Merrienboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/a:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tako, A. A., & Kotiadis, K. (2015). PartiSim: A multi-methodology framework to support facilitated simulation modelling in healthcare. *European Journal of Operational Research, 244*(2), 555–564. https://doi.org/10.1016/j.ejor.2015.01.046

Tedx Talks. (2023, November 2). *How failure drives learning| Manu Kapur| TEDxHSGSalon* [Video]. YouTube. https://www.youtube.com/watch?v=hv952Z0lfsI

Ten Berge, T., & Van Hezewijk, R. (1999). Procedural and declarative knowledge: An evolutionary perspective. *Theory & Psychology*, *9*(5), 605–624. https://doi.org/10.1177/0959354399095002

Tracy, M., Cerdá, M., Keyes, K. M. (2018). Agent-based modeling in public health: Current applications and future directions. *Annual Review of Public Health, 39*, 77–94. https://doi.org/10.1146/annurev-publhealth-040617-014317

Tonidandel, S., & LeBreton, J. M. (2013). Beyond step-down analysis: A new test for decomposing the importance of dependent variables in MANOVA. *Journal of Applied Psychology*, *98*(3), 469. https://doi.org/10.1037/a0032001

Turnbull, L., Hütt, M. T., Ioannides, A. A., Kininmonth, S., Poeppl, R., Tockner, K., ... & Parsons, A. J. (2018). Connectivity and complex systems: learning from a multi-disciplinary perspective. *Applied Network Science*, *3*, 1-49. https://doi.org/10.1007/s41109-018-0067-2

Uskola, A., & Puig, B. (2023). Development of Systems and Futures Thinking Skills by Primary Pre-service Teachers for Addressing Epidemics. *Research in*

Science Education (Australasian Science Education Research Association), 53(4), 741–757. https://doi.org/10.1007/s11165-023-10097-7

van Mil, M. H. W., Boerwinkel, D. J., & Waarlo, A. J. (2013). Modeling molecular mechanisms: A framework of scientific reasoning to construct molecular-level explanations for cellular behavior. *Science & Education, 22*(1), 93–118. https://doi.org/10.1007/s11191-011-9379-7

Vázquez-Serrano, J. I., Cárdenas-Barrón, L. E., Vicencio-Ortiz, J. C., Matis, T., Gaitán-Mercado, C. M., & Peimbert-García, R. E. (2024). Hybrid optimization and discrete-event simulation model to reduce waiting times in a primary health center. *Expert Systems With Applications*, *238*(Pt. B), Article 121920. https://doi.org/10.1016/j.eswa.2023.121920

Vázquez-Serrano, J. I., Peimbert-García, R. E., & Cárdenas-Barrón, L. E. (2021). Discrete-event simulation modeling in healthcare: A comprehensive review. *International Journal of Environmental Research and Public Health*, *18*(22), Article 12262. https://doi.org/10.3390/ijerph182212262

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic. *Family Medicine*, *37*(5), 360–363. https://studylib.net/doc/18216146/understanding-interobserver-agreement--the-kappa-statistic

Vosloo, J.-B., & Lemos, V. (2021, November 24). Learn java for AnyLogic! *AnyLogic Blog.* https://www.anylogic.com/blog/learn-java-for-anylogic/

Wanelik, K. M., Begon, M., Fenton, A., Norman, R. A., & Beldomenico, P. M. (2023). Positive feedback loops exacerbate the influence of superspreaders in disease transmission. *iScience*, *26*(5), Article 106618. https://doi.org/10.1016/j.isci.2023.106618

Weiss, I. R., Pasley, J. D., Smith, P. S., Banilower, E. R., & Heck, D. J. (2003). *Looking inside the classroom: A study of mathematics and science education in the United States.* Horizon Research Inc. https://horizon-research.com/insidetheclassroom/reports/looking/

Werler, M. M., Stuver, S. O., Healey, M. A., & LaMorte, W. W. (2019). The future of teaching epidemiology. *American Journal of Epidemiology, 188*(5), 825–829. https://doi.org/10.1093/aje/kwz039

Wilder-Smith, A., Gubler, D. J., Weaver, S. C., Monath, T. P., Heymann, D. L., & Scott, T. W. (2017). Epidemic arboviral diseases: Priorities for research and public health. *The Lancet Infectious Diseases, 17*(3), e101–e106. https://doi.org/10.1016/S1473-3099(16)30518-7

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology, 8*(1), 3–19. https://doi.org/10.1023/A:1009421303064

Wilensky, U. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. MIT Press.

Wilensky, U., & Reisman, K. (2006). Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theories—An embodied modeling approach. *Cognition and Instruction, 24*(2), 171–209. https://doi.org/10.1207/s1532690xci2402_1

Willem, L., Stijven, S., Tijskens, E., Beutels, P., Hens, N., & Broeckhove, J. (2015). Optimizing agent-based transmission models for infectious diseases. *BMC Bioinformatics, 16*, Article 183. http://dx.doi.org/10.1186/s12859-015-0612-2

Williams, R., Hosseinichimeh, N., Majumdar, A., & Ghaffarzadegan, N. (2023). *Epidemic modeling with generative agents.* arXiv. https://doi.org/10.48550/arXiv.2307.04986

Wood, D. F. (2003). Problem based learning. *BMJ*, *326*(7384), 328–330. https://doi.org/10.1136/bmj.326.7384.328

Woods, N. N., Brooks, L. R., & Norman, G. R. (2007). It all make sense: biomedical knowledge, causal connections and memory in the novice diagnostician. *Advances in Health Sciences Education: Theory and Practice*, *12*(4), 405–415. https://doi.org/10.1007/s10459-006-9055-x

Wu, B. (2024a). Decomposing the Cognition-influential Aspects of Instructional Designs: A Systematic Review of Induced Causal Relationships by the Constructional Components of the Direct Instruction and Productive Failure Designs. *Advances in Social Science and Culture*, *6* (5). 100-111. https://doi.org/10.22158/assc.v6n5p100

Wu, B. (2024b). The Competition between Productive Failure and Direct Instruction Is not Determinable: A Review of the Deviations and Reasons. *World Journal of Educational Research*, *11*(5). https://doi.org/10.22158/wjer.v11n5p172

Xia, S., Zhou, X. N., & Liu, J. (2017). Systems thinking in combating infectious diseases. *Infectious Diseases of Poverty*, *6*(1), Article 144. https://doi.org/10.1186/s40249-017-0339-6

Yang, Y., Atkinson, P. M., & Ettema, D. (2011). Analysis of CDC social control measures using an agent-based simulation of an influenza epidemic in a city. *BMC Infectious Diseases*, *11*, Article 199. https://doi.org/10.1186/1471-2334-11-199

Ylitalo, K. R., & Meyer, A. R. (2018). Bringing service into science: Community-engaged service-learning for undergraduate and graduate epidemiology students. *Pedagogy in Health Promotion*, *5*(2), 89–98. https://doi.org/10.1177/2373379918794970

Yoon, S. A., Chinn, C., Noushad, N., Richman, T., Hussain-Abidi, H., Hunkar, K., Cottone, A., Katz, J., Mitkus, E., & Wendel, D. (2023). Seven design principles for teaching complex socioscientific issues: the design of a complex systems agent-based disease epidemic model and the application of epistemic practices in high school biology. *Frontiers in Education*, *8,* 1210153.

https://doi.org/10.3389/feduc.2023.1210153

Zhao, H. (2024). Based on Productive Failure Theory Apply in STEM Learning. *Journal of Education and Culture Studies*, *8*(4), 86-102. https://doi.org/10.22158/jecs.v8n4p86

Appendix A: Ethics Approval Letters

THE UNIVERSITY OF SYDNEY

Research Integrity & Ethics Administration
HUMAN RESEARCH ETHICS COMMITTEE

Thursday, 9 March 2023

Prof Michael Joseph Jacobson
School of Education and Social Work Research Operations; Faculty of Arts and Social Sciences
Email: michael.jacobson@sydney.edu.au

Dear Michael Joseph,

The University of Sydney Human Research Ethics Committee (HREC) has considered your application.

I am pleased to inform you that after consideration of your response, your project has been approved.

Details of the approval are as follows:

Project No.:	**2022/583**
Project Title:	**Productive Failure and learning about epidemics and complex systems in medical education**
Authorised Personnel:	**Jacobson Michael Joseph; AlHinai Raeda;**
Approval Period:	**09/03/2023 to 09/03/2027**
First Annual Report Due:	**09/03/2024**

Documents Approved:

Date Uploaded	Version Number	Document Name
23/02/2023		Participant Information Statement
23/02/2023		Participant Consent Form
23/02/2023		Email Invitation to Participants
23/02/2023		Email Invitation to Instructors
22/11/2022		Letter to SQU
17/11/2022		Background Questions
17/11/2022		Experimental Group Modules
17/11/2022		Updated Methodology
17/11/2022		Control Group Modules
14/11/2022		Email Permission to the Assistant Dean
14/11/2022		Safety Protocol
11/07/2022		Interview Questions
11/07/2022		Pretest & Posttest
11/07/2022		Email Permission to the Dean

Special Condition/s of Approval

- In future, please provide a tracked-changes and clean version of any amended study documents.

Condition/s of Approval

- Research must be conducted according to the approved proposal.
- An annual progress report must be submitted to the Ethics Office on or before the anniversary of approval and on completion of the project.
- You must report as soon as practicable anything that might warrant review of ethical approval of the project including:
  - Serious or unexpected adverse events (which should be reported within 72 hours).
  - Unforeseen events that might affect continued ethical acceptability of the project.

Research Integrity & Ethics Administration
Research Portfolio
Level 3, F23 Administration Building
The University of Sydney
NSW 2006 Australia

T +61 2 9036 9161
E human.ethics@sydney.edu.au
W sydney.edu.au/ethics

ABN 15 211 513 464
CRICOS 00026A

THE UNIVERSITY OF
SYDNEY

- Any changes to the proposal must be approved prior to their implementation (except where an amendment is undertaken to eliminate *immediate* risk to participants).
- Personnel working on this project must be sufficiently qualified by education, training and experience for their role, or adequately supervised. Changes to personnel must be reported and approved.
- Personnel must disclose any actual or potential conflicts of interest, including any financial or other interest or affiliation, as relevant to this project.
- Data and primary materials must be retained and stored in accordance with the relevant legislation and University guidelines.
- Ethics approval is dependent upon ongoing compliance of the research with the *National Statement on Ethical Conduct in Human Research*, the *Australian Code for the Responsible Conduct of Research*, applicable legal requirements, and with University policies, procedures and governance requirements.
- The Ethics Office may conduct audits on approved projects.
- The Chief Investigator has ultimate responsibility for the conduct of the research and is responsible for ensuring all others involved will conduct the research in accordance with the above.

This letter constitutes ethical approval only.

Please contact the Ethics Office should you require further information or clarification.

Sincerely,

[redacted]

Associate Professor Haryana Dillon
Chair
Human Research Ethics Committee (HREC 3)

The University of Sydney of Sydney HRECs are constituted and operate in accordance with the National Health and Medical Research Council's (NHMRC) National Statement on Ethical Conduct in Human Research (2018) and the NHMRC's Australian Code for the Responsible Conduct of Research (2018)

Sultan Qaboos University
COLLEGE OF MEDICINE
& HEALTH SCIENCES

جامعة السلطان قابوس
كليـة الطب
والعلـوم الصحيــة

REF. NO. SQU-EC/ 136/2023
MREC #3038

TO: Dr. Hana Al Sumri
Principal Investigator, Department of Family Medicine & Public Health
Sultan Qaboos University Hospital

SUBJECT: *Approval of the Research Project*
"Productive Failure and Learning about Epidemics and Complex Systems in Medical Education"

DATE: 8th June 2023

I am pleased to inform you that the above mentioned research project submitted to the Medical Research Ethics Committee (MREC), College of Medicine and Health Sciences, Sultan Qaboos University for re-consideration and approval was discussed.

The last modifications received by the Committee on 8th June 2023 in response to the comments raised by during the 1st June 2023 meeting were found to be satisfactory.

The Committee has considered your research project acceptable and therefore **approval is granted.**

I wish you a productive study with your research work.

With kind regards,

Dr. Murtadha Al Khabori
Chairman, Medical Research Ethics Committee
College of Medicine & Health Sciences
Sultan Qaboos University

cc: Prof. Mohammed Al Zaabi, Asst. Dean, Postgraduate and Research Studies, COM&HS, SQU

P.O. Box: 35
Al-Khodh - Sultanate of Oman
Postal Code 123
Telephone: (+968) 24141103 Telefax: (+968) 24413419

صندوق البريد: ٣٥
الخوض - سلطنة عُمان
الرمز البريدي: ١٢٣
هاتف: ٢٤١٤١١٠٣ (+٩٦٨) فاكس: ٢٤٤١٣٤١٩ (+٩٦٨)

Appendix B: Participant Consent Form

THE UNIVERSITY OF SYDNEY

School of Education and Social Work

ABN 15 211 513 464

PROFESSOR MICHAEL J. JACOBSON
Professor and Chair of Education,
Centre for Research on Learning and Innovation

The University of Sydney
NSW 2006 AUSTRALIA
Email: michael.jacobson@sydney.edu.au
Web: http://www.sydney.edu.au/

Productive Failure and Learning about Complex Systems and Epidemics in Medical Education

PARTICIPANT CONSENT FORM

I,... [PRINT NAME], agree to take part in this research study.

In giving my consent I state that:

- I understand the purpose of the study, what I will be asked to do, and any risks/benefits involved.
- I have read the Participant Information Statement and have been able to discuss my involvement in the study with the researchers if I wished to do so.
- The researchers have answered any questions that I had about the study, and I am happy with the answers.
- I understand that being in this study is completely voluntary and I do not have to take part. My decision whether to be in the study will not affect my relationship with the researchers or anyone else at the University of Sydney, Sultan Qaboos University or any other individuals or organisation related to this study, now or in the future.
- I understand that I can withdraw from the study at any time.
- I understand that I may stop the interview at any time if I do not wish to continue. I also understand that I may refuse to answer any questions I do not wish to answer.

- I understand that personal information about me that is collected over the course of this project will be stored securely and will only be used for purposes that I have agreed to. I understand that information about me will only be told to others with my permission, except as required by law.

- I understand that the results of this study may be published, but these publications will not contain my name or any identifiable information about me.

  I consent to:

  - **Audio-recording** YES NO

I would like to review my interview transcripts YES NO

I would like to receive feedback about the overall results of this study YES NO

If you answered **YES**, please indicate your preferred form of feedback and address:

Email: ____

....
Signature

....
PRINT name

....
Date

Appendix C: Participant Information Sheet

THE UNIVERSITY OF SYDNEY

School of Education and Social Work

ABN 15 211 513 464

MICHAEL J. JACOBSON
Professor and Chair of Education,
Centre for Research on Learning and Innovation

The University of Sydney
NSW 2006 AUSTRALIA
Email: michael.jacobson@sydney.edu.au
Web: http://www.sydney.edu.au/

Productive Failure and Learning about Complex Systems and Epidemics in Medical Education

PARTICIPANT INFORMATION STATEMENT

1. What is this study about?

You are invited to take part in a research study about epidemics concepts with the use of computer models. Your course instructor will introduce the research topic, but you will work with the study materials independently on the Moodle platform for 3 sessions.

This Participant Information Statement tells you about the research study. Knowing what is involved will help you decide if you want to take part in the research. Please read this sheet carefully and ask questions about anything that you don't understand or want to know more about.

Your participation in this research study is voluntary.

By giving your consent to take part in this study you are telling us that you:

- Understand what you have read.
- Agree to take part in the research study as outlined below.
- Agree to the use of your personal information as described.

You will be given a copy of this Participant Information Statement to keep.

2. Who is running the study?

The study is being carried out by Raeda Al Hinai, PhD Student, Sydney School of Education and Social Work, The University of Sydney.
Raeda Al Hinai is conducting this study as the basis for the degree of PhD at The University of Sydney. This research takes place under the supervision of Professor Michael Jacobson from the Sydney School of Education and Social Work, University of Sydney.

3. What will the study involve for me?

You will use your Moodle account to access the study materials, which mainly involve watching short videos about epidemics diseases, using computer models, and answering problem solving questions. There will also be optional focus group interviews with groups of 6 to 8 participants.

4. How much of my time will the study take?

There will be three sessions of two hours each. At the beginning of the first session, there will be a pretest and at the end of the last session, there will be a post-test. The optional focus groups will each take approximately 30 minutes.

5. Who can take part in the study?

Any medical student who enrols to study a course entitled, " Hospital and Community Attachment" can participate in this research project.

6. Do I have to be in the study? Can I withdraw from the study once I've started?

This study is completely voluntary, and you do not have to take part. Your decision whether to participate will not affect course grade or your current or future relationships with the individuals. If you decide to take part in the study and then change your mind later, you are free to withdraw at any time. If you withdraw, then all the information you provided during the study will be deleted.

7. Are there any risks or costs associated with being in the study?

Aside from giving up your time we do not expect that there will be any risks or costs associated with taking part in this study.

8. Are there any benefits associated with being in the study?

As you are an undergraduate medical student, it is anticipated that this study has the potential to enhance your understanding of the epidemics ideas, although there is no guarantee that you will receive any direct benefits from being in this study. Also, you may find the use of computer models to be an interesting way to receive supplemental information related to your overall medical degree program.

9. What will happen to information about me that is collected during the study?

By providing your consent, you are agreeing to us collecting personal information about you for the purposes of this research study. Your information will only be used for the purposes outlined in this Participant Information Statement, unless you consent otherwise.

Your information will be stored securely, and your identity/information will be kept strictly confidential, except as required by law. Study findings may be published, but you will not be individually identifiable in these publications.

10. Can I tell other people about the study?

Yes, you are welcome to tell other people about the study.

11. What if I would like further information about the study?

When you have read this information, Ms Raeda Al Hinai will be available to discuss it with you further and answer any questions you may have. If you would like to know more at any stage during the study, please feel free to contact *Ms Raeda Al Hinai*, student researcher, email: ralh2617@uni.sydney.edu.au, telephone: +96892926799 or Professor Michael Jacobson, supervisor, email: michael.jacobson@sydney.edu.au.

12. Will I be told the results of the study?

You have a right to receive feedback about the overall results of this study. You can tell us that you wish to receive feedback by *ticking the relevant box on the consent form.* This feedback will be in the form of a one-page lay summary. You will receive this feedback after the study is finished.

13. What if I have a complaint or any concerns about the study?

Research involving humans in Australia is reviewed by an independent group of people called a Human Research Ethics Committee (HREC). The ethical aspects of this study have been approved by the HREC of the University of Sydney [2021/243]. As part of this process, we have agreed to carry out the study according to the *National Statement on Ethical Conduct in Human Research (2007).* This statement has been developed to protect people who agree to take part in research studies.

If you are concerned about the way this study is being conducted or you wish to make a complaint to someone independent from the study, please contact the university using the details outlined below. Please quote the study title and protocol number.

The Manager, Ethics Administration, University of Sydney:

- **Telephone:** +61 2 8627 8176
- **Email:** human.ethics@sydney.edu.au
- **Fax:** +61 2 8627 8177 (Facsimile)

In addition, a local contact for any complaint or concern is provided below:

Mohammed Abdullah AlZaabi, Assistant Dean for Postgraduate Studies and Research, College of Medicine and Health sciences, Sultan Qaboos University

- **Telephone:** +968 24143431
  Email: zaadi@squ.edu.om

Appendix D: Experimental Group Sessions Materials

(Provided to students after the Pretest)

Introduction

The research study will be conducted online in Moodle platform asynchronously. You can use your university account to access the platform and the content is available within your course webpage. Please follow the instructions below to complete the three sessions. During these sessions, you will be using computer modelling software, called Anylogic, to explore the three main aspects of epidemic diseases:

1. The Dynamics of Epidemic Diseases
   - How the dynamics of epidemic diseases can be applied in a population through SEIRS
2. Malaria as An Epidemic Disease
   - How Malaria can transmit in a population
   - How preventive measures can affect the prevalence of Malaria
3. COVID-19 as An Epidemic Disease
   - How COVID-19 can transmit in a population
   - How the global response related to coronavirus can prevent the prevalence of the disease.

The models that you will be exploring in both fields will help you deeply understand the behaviour of the virus, its agents' interactions that arise at the micro and macro level of the system, and the mitigation policies used to prevent the spread of such diseases.

Each session will consist of 4 separate phases and technically, you will not be able to move to the second phase unless you have completed the previous phase. It is very important to follow the sequence of the phases of each day session:

- Phase 1: You will be asked to activate your knowledge and explore a computer model to find answers to questions related to epidemics ideas. You will also be asked to explain your answers;
- Phase 2: You will be asked to explore a model and find answers to questions related to complexity ideas. You will also be asked to explain your answers;
- Phase 3: You will watch a guided learning video to understand the key concepts you may have encountered in phases 1 and 2;
- Phase 4: After consolidation, you will be asked to solve a real-world based problem that combines the targeted concepts of both complex systems and epidemics.

During the sessions, you will be asked to work on problems using two different models each day of the study. For all of the problems you work on with the different models, you should come up with as many ideas as you can for how you might solve the problems. There are multiple ways that the problems may be solved, so do your best with coming up with your own solution ideas.

Session 1: The Dynamics of Epidemics

Phase 1 – Epidemics Model (SEIRS) – Susceptible, Exposed, Infected, Recovered

The model presents the following variables:

Latency: time it takes from the moment the person gets the virus until it becomes contagious

- Infectious: amount of time the individual stays infectious
- Immune: amount of time the person stays immune
- Infection probability: the chance for someone to infect someone else once a contact has been made
- Contacts per day: the number of times an individual gets is in contact with another person/day.

Note: in the models proposed, people do not die.

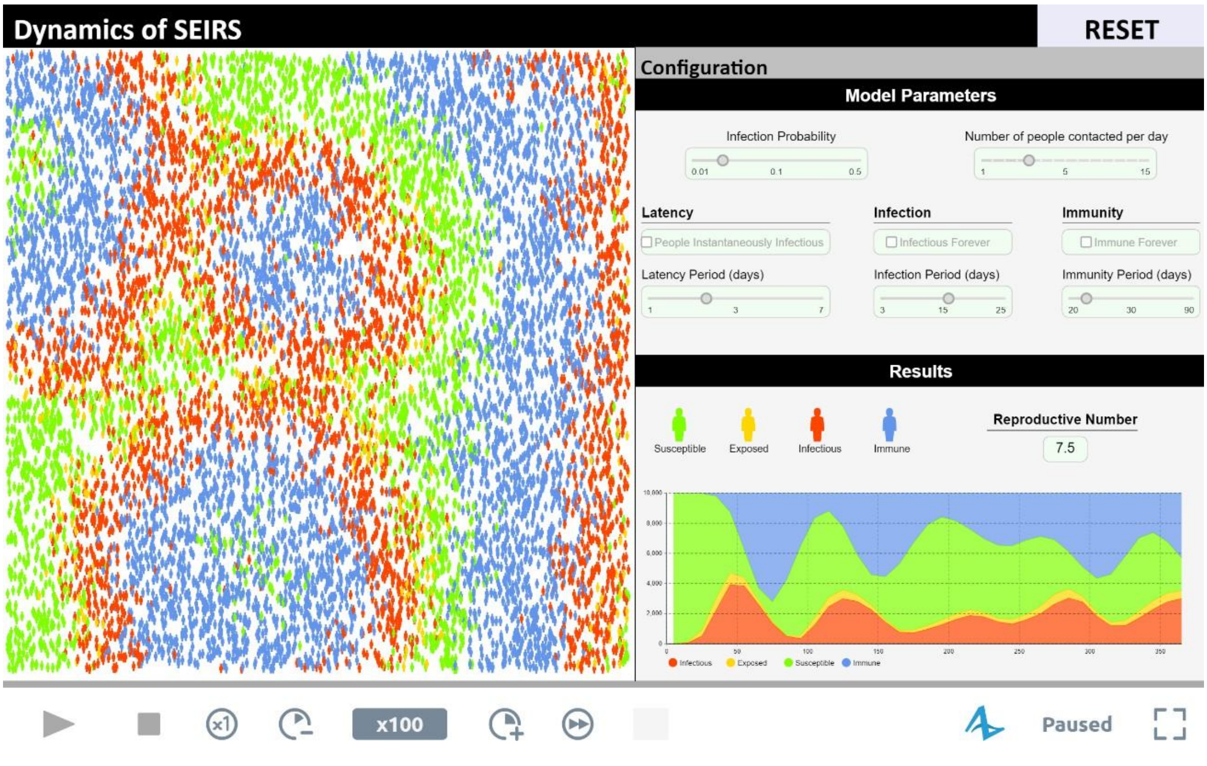


Figure B1

A Screenshot of the SEIRS Model Interface

Question:

- How is it that the population constantly gets infected and the disease continues to spread over time? Write down as many ideas as you can come up with, that might answer this question. It is not important to identify a "right" answer.

Phase 2 – Forest Fire Model

This model consists of a forest with trees of a certain density. A fire has started on the west section of the forest and the fire will spread towards the east. This is a forest in which trees have certain fire-resistant characteristics. The parameters determine how dense the forest is (how close the trees are to each other) and the fire resistance (a resistant tree will not be burned when the fire comes).

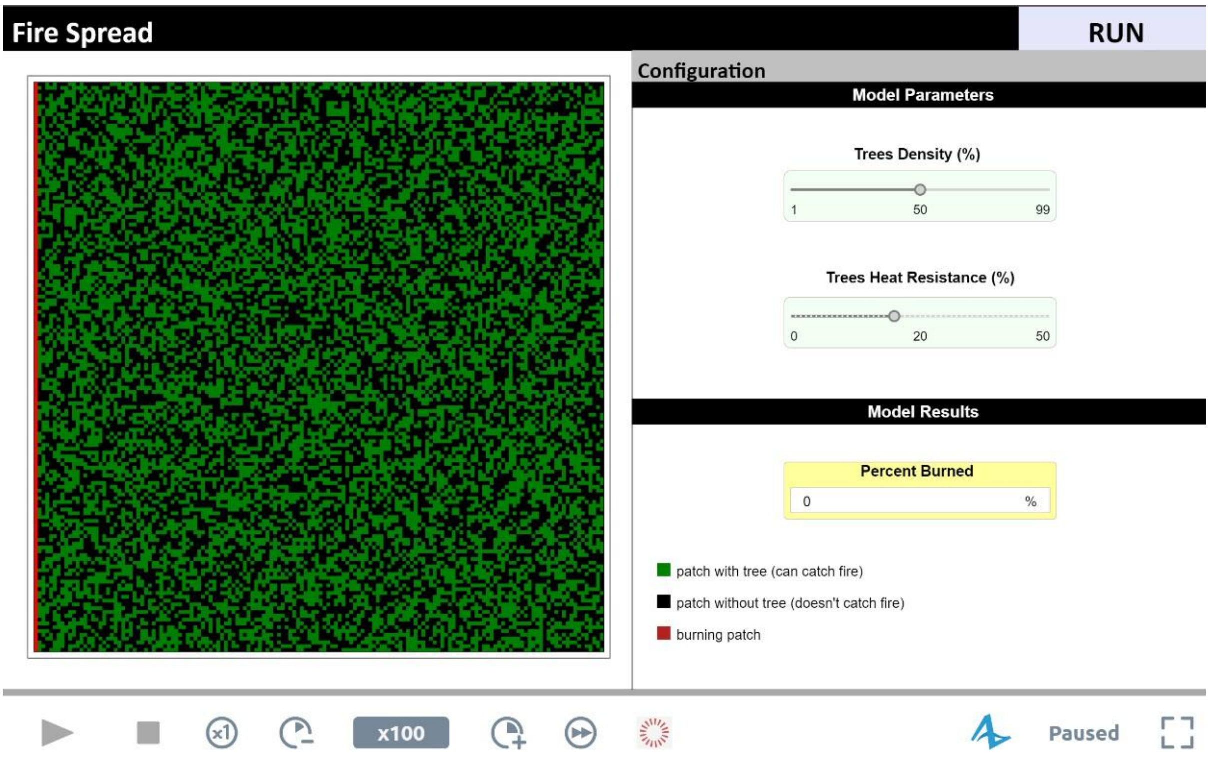


Figure B2

A Screenshot of the Forest Fire Model Interface

Question:

The forest world is built in the simulation with pixels. Each pixel (or patch) represents either a tree, fire, or empty space. When you watch the simulation run, it looks like the fire moves, nevertheless, none of these pixels (or patches) actually move. Write as many ideas as you can come up with, that might answer why the fire never actually moves. It is not important to identify the "right" answer.

Phase 3 – Discussion

After exploration, you will watch a guided learning video that includes some of the anticipated answers you have thought about and explain the key concepts of epidemics and complex systems you have used in the previous phases:

- Dynamic equilibrium
- Dynamics of SEIRS, latency, prevalence, infectivity, mortality

Phase 4 – Application of Epidemics and Complex Systems Concepts

Questions:

- You are a scientist who works in the public health department of a Middle Eastern country. The Minister of Health has asked you to advise on the dangers associated with a highly infectious disease that has recently been identified, with unknown characteristics. Provide the Minister a summary of your assessment on how the disease is spread amongst the population depending on the disease characteristics. Please present your summary in general terms so that the public would understand.
- Write down what you think are the main similarities and differences between the two models you have looked at.

Session 2: Malaria as An Epidemic Disease

Phase 1 – Malaria Model

The parameters that are important to know are:

- Initial population: people present in the simulation
- Initial prevalence: how many people are infected at the start of the simulation
- Contact rate (mosquito): how many people are reached by mosquitos every day
- Hospital capacity
- ICU capacity
- Fraction of infected requiring hospital
- Fraction of infected requiring ICU

Note: The other parameters are similar to the ones seen in the previous day.

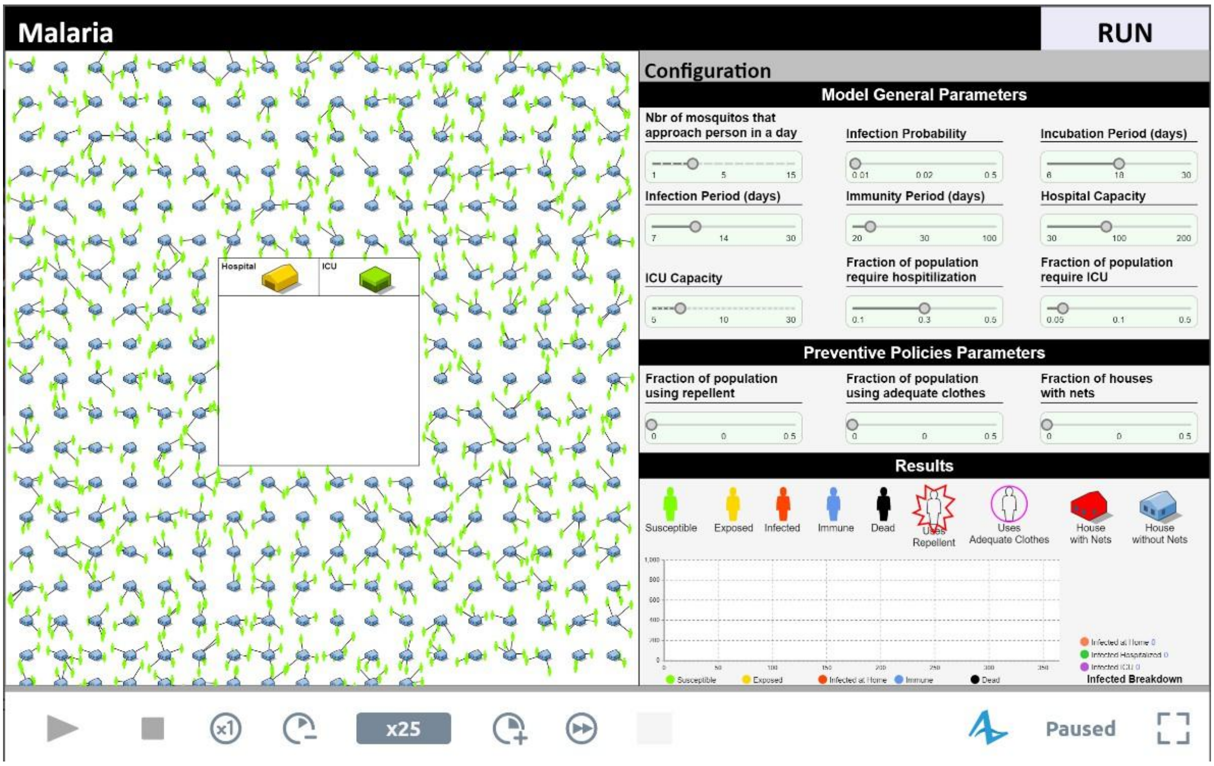


Figure B3

A Screenshot of the Malaria Model Interface

Question:

Explain how the disease spreads over time and the population's behaviour towards this. Write as many ideas as you can come up with, that might answer this question. It is not important to identify the "right" answer.

Phase 2 – Marketing Model

The Marketing model is presented with the following parameters:

Contact Rate: number of people someone might get in contact with during a day

- Adoption Fraction: the probability for someone to buy the product once they become aware of the brand
- Product Quality: a measure that determines how likely someone is to like the product and become loyal to the brand
- Product Lifetime: how much time it takes for someone to need the product again after purchase.

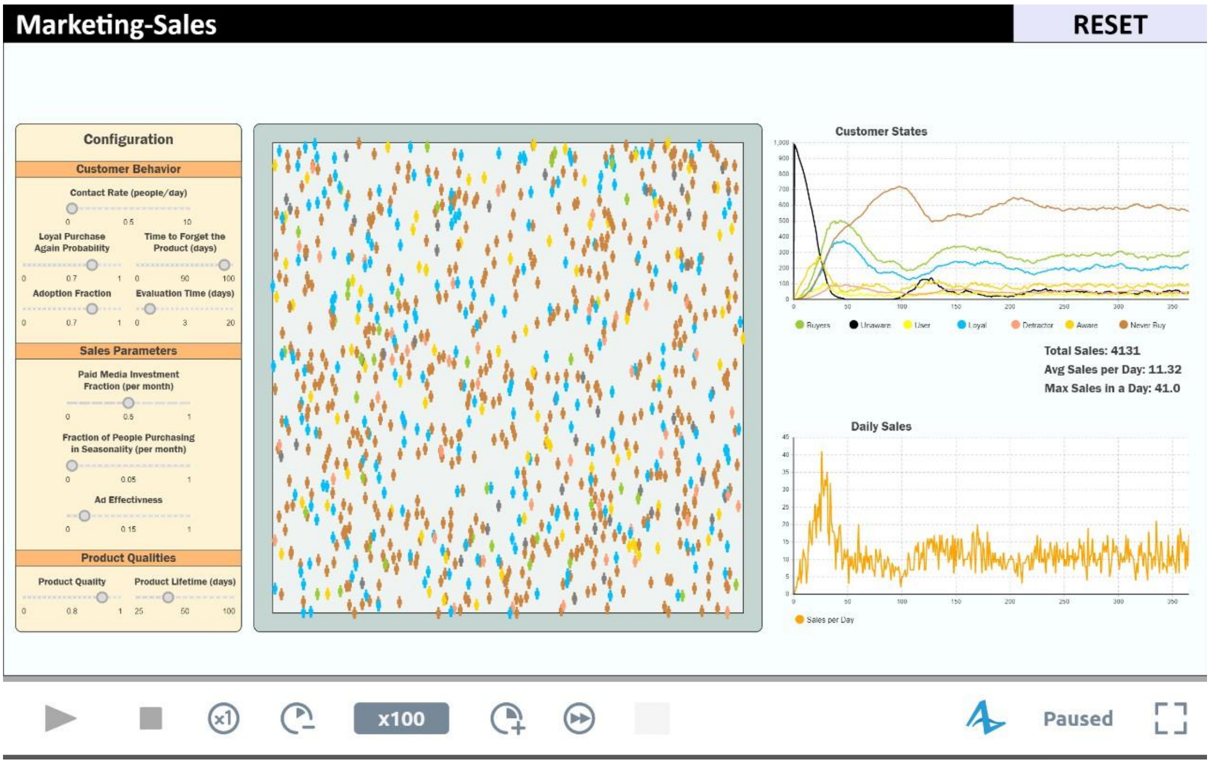


Figure B4

A Screenshot of the Marketing Model Interface

Question:

- Explain the people choices over time and the factors that affect the population's behaviour. Write as many ideas as you can come up with, that might answer this question. It is not important to identify the "right" answer.

Phase 3 – Discussion

After exploration, you will watch a guided learning video that includes some of the anticipated answers you have thought about and explain the key concepts of epidemics and complex systems you have used in the previous phases:

- Tipping Points
- Animal-Human Transmission, Mitigation Policies

Phase 4 – Application of Policies in Malaria Model

The same model is used as above, but giving more emphasis on mitigation:

- Fraction of people using repellent
- Fraction of people using appropriate clothes
- Fraction of people using mosquito nets

Questions:

- You are a researcher working at the local university and you are asked for recommendation policies to implement in response to the Malaria situation based on your current research findings. Think about the recommendations you can make and their implication in the way the population will react compared to the results in reducing the spread of this disease. Explain your reasoning for a non-technical audience, using epidemics and complex systems concepts as to why your policy recommendation will be effective in the short and/or long term.
- Please explain how the concept of Tipping Points applies to the two models you have investigated.

Session 3: COVID-19 as an Epidemic Disease

Phase 1 – COVID-19 Model:

The same parameters previously studied are used in this model. The new one is fraction asymptomatic that adds two possible states to the infected person: symptomatic and asymptomatic.

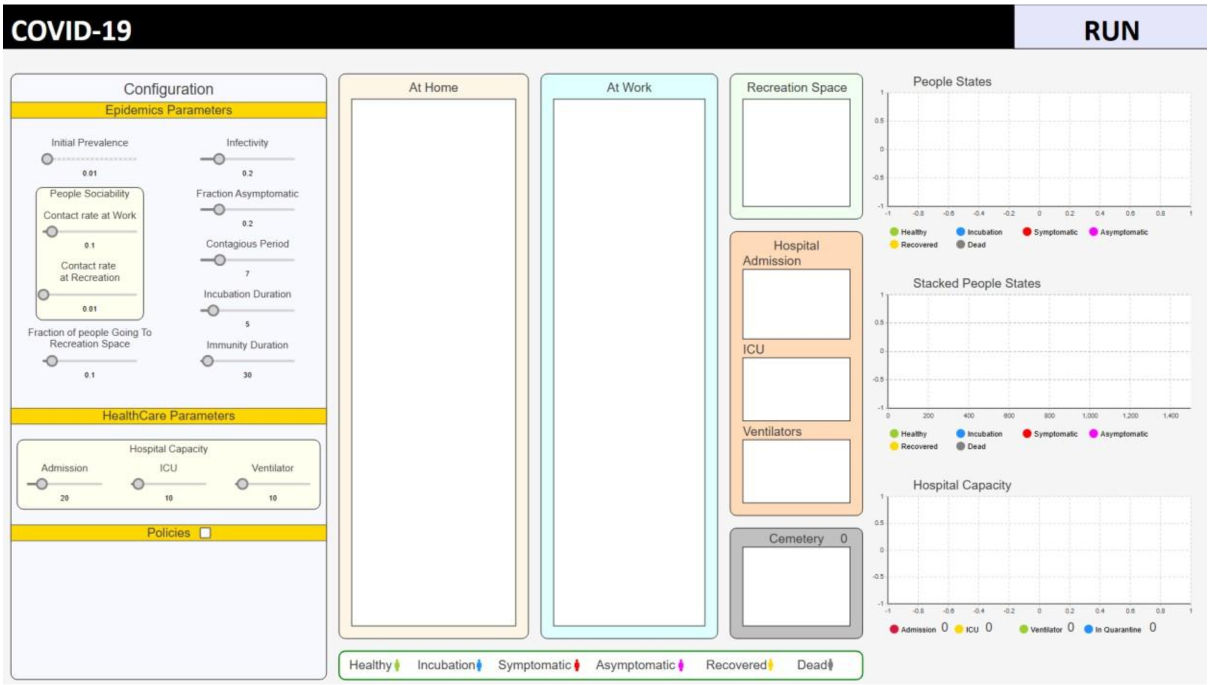


Figure B5

A Screenshot of the COVID-19 Model Interface

Question:

Explain how the disease spreads over time and the factors that affect the populations' behaviour? Write as many ideas as you can come up with, that might answer this question. It is not important to identify the "right" answer.

Phase 2 - Wolf-Sheep Predation Model - Dynamic Equilibrium

The model presents the following dynamic options:

- To use grass or not
- Initial number of sheep and wolves
- Grass regrowth time
- Sheep and sheep energy gain from food
- Reproduction rates of sheep and wolves

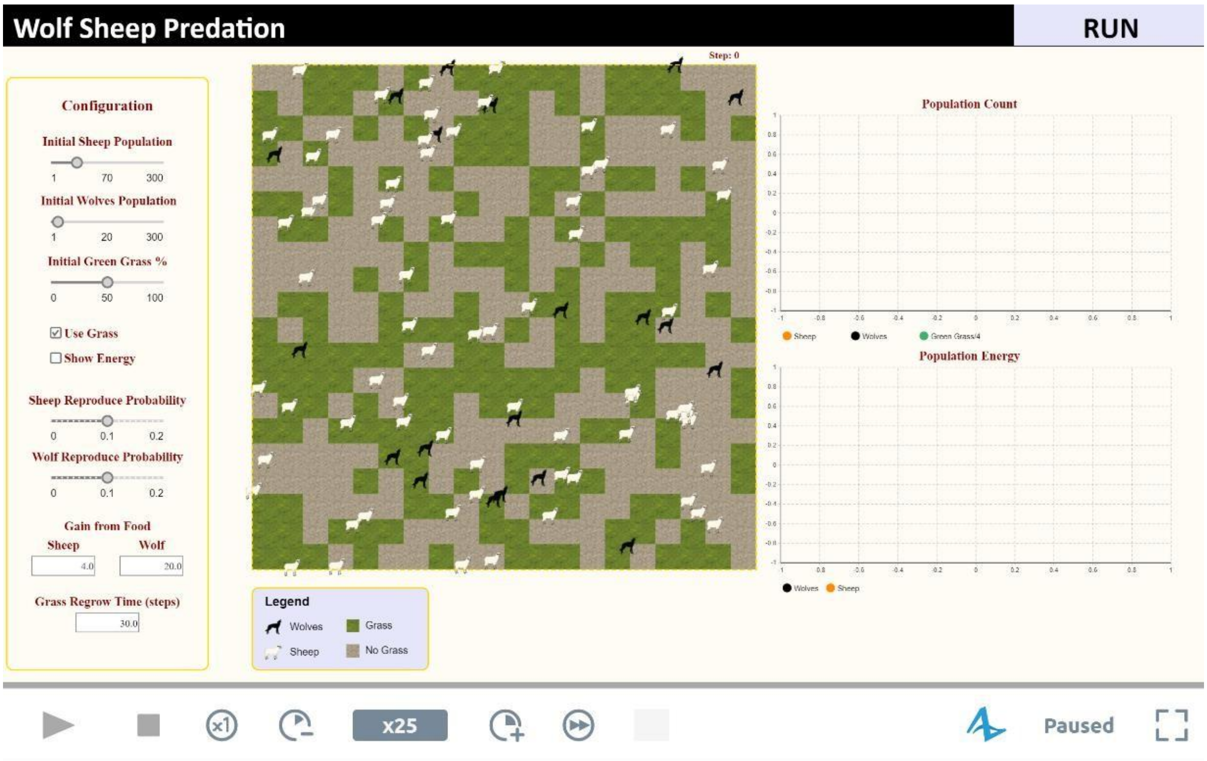


Figure B6

A Screenshot of the Wolf-Sheep Predation Model Interface

Questions:

- How is it that the sheep do not eat all the grass or that the wolves do not eat all of the sheep? Write down as many ideas as you can come up with, that might answer this question. It is not important to identify the “right” answer.
- How would you describe the type of relationship that you observe between grass, sheep, and wolves?

Phase 3 – Discussion

After exploration, you will watch a guided learning video that includes some of the anticipated answers you have thought about and explain the key concepts of epidemics and complex systems you have used in the previous phases:

- Emergent behaviour (micro and macro- systems' levels)
- Human-human transmission
- Mitigation policies

Phase 4 – Application of Policies in COVID-19 Model

The same model is used as above, but mitigation policies have been added:

- Fraction of people using masks
- Fraction of people in isolation
- Fraction of people following hygienic guidelines
- Vaccine

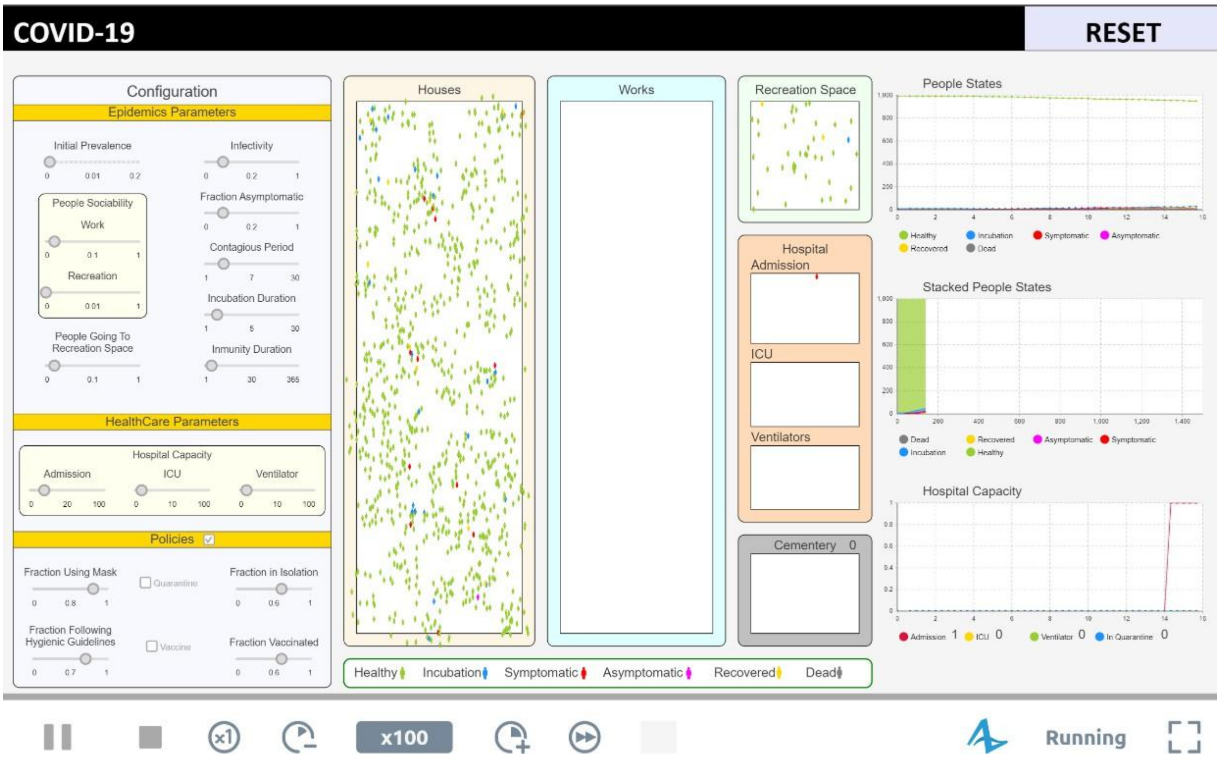


Figure B7

A Screenshot of Mitigation Policies in the COVID-19 Model

Questions:

- As a Director of the Inpatient ward at the local hospital, analyse the way hospitals respond during an increase in symptomatic populations. Propose some appropriate recommendations for the healthcare professionals to reduce the mortality in the population.
- Write down what you think are the main similarities and differences between the two models you have looked at.

Appendix E: Instructional Video Scripts for All Sessions

Session 1 Instructional Video Transcript

Introduction: Hello, and welcome, Medical Students. My name is Gina, and I am a lecturer in Australia and teach in both the MD program and the Population and Global Health undergraduate and postgraduate degrees. I will be here to share my knowledge about epidemics and complex systems. Let's start learning.

Concept Definition and Description:

Emergence is a central concept in complex systems research. It is generally defined as the formation of new properties or patterns at a system's macrolevel from self-organizing interactions of elements at a system's microlevel. A commonly experienced example of emergence is seen when a group of birds are flying together: there is often a single bird ahead of the other birds, who spread out in a "V" shape formation. Here the "microlevel elements" are the individual birds (who do not look like a "V", of course) and the "macrolevel property" is the emergence of the "V" shape flock of birds. Other examples of emergence include chemical equilibrium, where the microlevel interactions of molecular reactants over time show no macrolevel observable change in the properties of the system, and the microlevel interactions of burning trees spreading to other trees if they are close enough, resulting in a macrolevel forest first.

Section 1: Epidemics (SEIRS Model)

The SEIRS model is usually used in epidemiology to study how a disease is going to spread amongst the population. It is used when there is a new disease that we don't know a lot about its characteristics. We use the model to find out how long the disease will take to spread and to help understand how long it will take to end. There are also different versions of the model, such as SIR or SIRS. All these versions are not related to a specific type of disease but instead can be used to study any type of disease. Each letter of SEIRS stands for something different.

- S stands for susceptibility, which means the person is susceptible to the disease. The person can acquire the disease because they have not acquired it before and do not have immunity against it.
- E stands for exposure. During the time a person is exposed, they cannot infect others with the disease. There is a latency period, which is the length of period where the person has the disease but cannot transmit it to other people. This is also called the "incubation period", and it is used when describing infectious diseases.
- I stands for infected. It means the person has the disease and can infect others or transmit to others.
- R stands for recovered, which means the person is now totally recovered from the disease and is immune for a certain period.
- S stands for susceptible. When the person has passed the immunity period, they can get the disease again.

There are a number of parameters that you can control.

- 'Infection probability' and 'the number of people contacted per day' are the main factors that will specify how the disease will be spread. The 'number of people contacted per day' are the number of people that each person is going to come in contact with or be in the same area or meet with each day. 'Infection probability' is the probability of transmitting the disease to each one of these persons. For example, if the probability is 0.01 and the person will meet 5 people each day, then there is a 1% chance that the person will transmit the disease to each person that they meet.
- Latency. This is the period in which the person is exposed. This is the period that the person acquires the disease but cannot transmit it to others. You have the option in the model of making the person instantaneously infectious, which makes the latency

period equal to 0. If you choose to have the latency period as 0, when you press that button, it means you cannot control the latency period anymore, as it is forced to be 0.

- Infection (Period). This is the duration the person is still infected and can transmit the disease to other people. You can specify whether or not the person is infectious forever, which means that the person will always have the disease and will always be able to transmit it to other people. An example of this is when a person has HIV. HIV is the kind of disease you acquire and do not recover from; therefore, you have it 'forever' and will be able to transmit it to others indefinitely.
- Immunity. This is the length of period in which the person remains immune. So, after the person has recovered, they will remain immune for a certain period of time. During this period, the person cannot acquire the disease. Like the other factors, you can also specify that once the person acquires the disease and recovers, they can become immune forever. This means they will never acquire the disease again.

The chart results are visualised through a stack chart, where each colour corresponds to a specific state rather than its overall height. By referring to the legend, you can identify which colour corresponds to each state. The chart showcases the current status of the population, and it is important to interpret the values as differences between the states, not cumulative totals. By hovering over specific sections of the chart, you can observe the difference between that particular state and others. It is important to understand that the number of people in each state is represented by the width of the strip and not its total height. Let's take as an example the results shown in the figure below. At time t=50, we have around 4000 infectious people as we can read from the red graph. The number of exposed people (yellow) is equal to around 4500- 4000 = 500 (and not 4500). We can easily see that at that time, the number of infectious people or susceptible people is greater than the number of exposed people because of the widths of their relevant colour strips. The exact values are not as important as noticing the trends in the changes of the states together. Note that the sum of the number of people in each state at any time will always be equal to 10,000, which is the total number of people in the population in the model.

Parameters Set 1 Results:

These parameters show the typical behaviour of disease spread where we have waves that peak at certain points and then decrease again. The contact rate used here is 5 per day, which means that there is good probability of a person acquiring the disease; however, the probability is not high enough for the disease to spread very rapidly, and this is why we see this wave like structure. The disease spreads progressively, and thus there will always be infected people to transmit the disease again. We can also see from the animation the emergence patterns in the system that reflect the propagation trends of the disease. This means that we can see clear patterns in the arrangement of colours. The colours are not dispersed or arranged randomly but form organized shapes that can be visually distinguished. The narrow width of the colour strips or ripples shows the progressive rate of spread. At the micro level, the entities, which are people in this case, are randomly distributed at the beginning of the model, with each of these entities having certain characteristics that are also randomly assigned. At the macro level or the whole system level we were able to see the formation of patterns that cannot be explained or predicted from the entities' individual properties or behavioural interactions, and this is actually what distinguishes the phenomena of emergence.

Parameters Set 2 Results:

The increased contact rate to 9 per day in this parameter set gives a more severe jump in the number of infected people. Both the graph and the animation show how the disease spreads very quickly such that most people get infected around the same time and also get cured around the same time, so the infections will reach zero after the first spike and then there will be no one left with the disease to infect others. The wide strips in the emerged patterns from the animation show the speed of the spread in this case.

Parameters Set 3 Results:

Although the rate of contact in this parameter set is set back to 5 per day, the fact that the person becomes infectious forever yields a similar behaviour to the one seen in parameter set 2 in terms of the speed of spread of the disease. However, in this case the graph continues moving upwards (we notice a sort of an s-curve where we have a slow rate of increase at the beginning, then a faster increase rate in the middle, and finally the rate decreases again in the last upper part of the graph) until everyone gets infected. This is because, being infectious forever, a person will always have a probability of infecting any of his/her contacts. Here again we see wide strips in the animation; however, the dominant colour in this case is red since everyone eventually gets infected.

Section 2: Emergence as a Complexity Concept (Forest Fire Model)

Emergence refers to the phenomenon where complex patterns, behaviours, or properties arise from the interactions and collective behaviour of simpler components or entities within a system. The fire model is a typical model that is usually used to show this phenomenon. Forest fire models often capture the emergent behaviour of fire propagation and its complex patterns. There are different versions of this model where the behavioural rules and the starting conditions are altered, but all can clearly show the emergence as a concept.

Layout Description:

In the model, the forest layout is represented by small cells or patches that together constitute the forest. Each cell can be empty or can have a tree. Each cell has four neighbour cells that it can be affected by. These are the cells in the four main directions (north, south, east, west).

Rules of fire spread:

- Cells with trees can burn
- Empty cells can't burn
- A cell will burn if at least one neighbour is burning
- The fire always starts on the left edge of the forest

Parameters Description:

- Tree density: the percentage of cells that are trees (not empty)
- Trees heat resistance: the ability of the cells to resist the fire. For example, if the heat resistance is set to 30%, this means that the probability of a cell catching the fire from a neighbouring cell is reduced to 70%.

Results:

The results you will see in this model are simply the percentage of tree cells that are burned. However, even more important is the graphical results of the model that will show the emergent behaviour. The controlling factor in this model is the density of the trees. There seems to be a threshold at which the spread behaviour changes. This threshold is around 59% tree density. When the tree density is less than this, the fire cannot reach the other end of the forest.

Parameter Set 1 Results:

This is a case where the tree density is lower than the threshold (50%). We can see from the animation and the results that the fire was able to spread to a very small part of the forest area with a total of 4.7% of patches burned. The patches with no trees were able to block the spread of the fire.

Parameter Set 2 Results:

With a 70% tree density, we see a completely different behaviour where the fire easily moves through the forest to reach the most-right edge, with a total percentage burned of around 98%. During the simulation, we see from the animation how the system emerges to a wave like movement or spread of the fire. Although the individual patches never move at the micro level, the overall resultant behaviour of the system (macro level) conveys movement.

Parameter Set 3 Results:

This is the extreme case where we have a 99% tree density. Again, here we notice a new emergent behaviour. The system conveys movement but this time the movement patterns are much faster and smoother.

In parameter set 3, we see a behaviour similar to slow crashing waves that are gradually losing momentum, while in parameter set 2, we see something similar to a very strong wave that moves rapidly and reaches the shore with the same power. Even in parameter set 1, we see a behaviour that resembles small waves that get blocked by huge rocks that end the waves' lifetime. All these visual impressions clearly show the concept of emergence. It is important to note that the trees' heat resistance plays a mediating role in this model. In other words, having a 99% trees density with a 50% resistance, for example, will yield a similar behaviour to a 50% tree density with 0% heat resistance.

Session 2 Instructional Video Transcript

Introduction: Hello, and welcome, Medical Students. My name is Gina, and I am a lecturer in Australia who teaches in both the MD program and the Population and Global Health undergraduate and postgraduate degrees. I will be here to share my knowledge about epidemics and complex systems. Let's start learning.

Concept Definition and Description:

A tipping point is a threshold moment when a system dramatically or rapidly changes from one stable state to another. These shifts in state usually occur as a response to a small change or the accumulation of previous small changes. Common examples of tipping points (sometimes called "phase transitions") include liquid water changing to ice at zero degrees Celsius or a song from a new artist suddenly becoming a "hit" that everyone listens to. Small changes in a system may be described as "linear" changes, but the rapid change to a different stable state represents a "non-linear" change. In many biological and social systems, it is difficult to predict when changes go from being linear to non-linear, that is, when a tipping point might occur.

Section 1: Epidemics (Malaria Model)

Initially, compared to the previous session (SEIRS), some parameters are not available to manipulate while playing with the model such as instantaneously infectious, immune forever, and infectious forever. WHY? Because here we are simulating a known disease whose properties we know. Malaria is not instantaneously infectious, as there is an incubation period of between 6 and 30 days. Also, people who get malaria will recover and will not be infectious forever.

We have introduced some new concepts in this model like death, hospitalisation, and ICU. When people are infected with Malaria disease, there is a certain probability that they will be Hospitalized, and of those that require hospitalisation, there are some who will need to go to the ICU. Because of this, the hospital will have a certain capacity and the ICU will have a certain capacity. In the animation, the hospital or the ICU areas will become red and indicate the word "FULL" when there are no current spaces. If a person requires hospitalisation and there is no space available there, there will be a certain period after which the person may die, and the same scenario occurs for the ICU.

As you can see in the animation window, there are some houses and people connected to their houses with lines to show whether they are in or out. When you run the model, you will notice the people are distributed in different boxes; they could be at home, outside, hospitalised or in ICU. So, the person who is inside the house is less at risk of being exposed to mosquitoes. The mosquitoes are the ones transmitting the disease in this scenario. People cannot transmit this disease to each other. It is not person to person. The only way the disease can be transmitted is if a mosquito that holds the disease bites someone who doesn't; there is then a chance that the disease will be transmitted to the person.

First, let's have a closer look at the general parameters of this model:

- Number of mosquitoes that approach a person in a day.
- Infection probability. This is similar to the one in SEIRS but instead, we are talking here about contact with mosquitoes.
- Incubation Period means the person has the disease, but they do not show any symptoms. We can't say "they have it but cannot transmit it" because they cannot transmit it anyways.
- Infection period is the time in which the person will have the disease and will show symptoms of the disease.
- Immunity period, which is the duration that the person will stay immune. Note that in the case of malaria, a person might not necessarily become immune directly after the first infection. In most cases, it might take several infections for the person to become immune. In the case of malaria, when the person is infected, they may have symptoms, but they may not require hospitalisation depending on the severity of the disease.
- Hospital Capacity/ Fraction of population that requires hospitalisations. So, you can determine the percentage of people who are infected and will require hospitalisation.
- ICU Capacity/ Fraction of population that requires ICU. The number of beds the ICU has or the number of people it can take. This unit usually has a smaller number of beds or rooms to take patients.

Keep in mind, if a person needs to find a hospital bed, and a bed is not available, they can only wait for a short duration; if this time passes and they do not find a bed in the hospital or

ICU, they will die.

Second, let's move to the preventive policies parameters.

These policies are considered by the government or even by a person to reduce their chance of being infected by malaria. We have chosen three policies:

1. Using a repellent, a type of cream, or a spray will help to prevent a mosquito biting the person and transmitting malaria. This is why using the repellent reduces the infection probability. It is represented in the model by a person with a red shape around him (Check legend). This does not reduce the number of mosquitoes that will be present to bite you but will reduce the probability of being bitten by a mosquito.
2. Using adequate clothing cover, including long sleeves to cover your arms and pants instead of shorts to reduce the area of exposure. This is represented in the model by a person with the purple circle around him. This policy can become even more stringent in cases when the rate of malaria is high so that people even wear items on their heads or gloves on their hands when they are outside.
3. Using nets inside houses. Different types of nets can be put inside houses to cover the beds. When people are sleeping, the mosquitoes cannot go through the nets. For example, if you scroll the parameter and select 0.5 as a fraction, that means 50% of the population are using that type of policy. When running the model, you will notice, in the animation window, the houses with nets will be changed to red colour while the houses without nets will stay blue.

Parameter Set 1 Results:

The first set of parameters is basically the default ones we get when we run the model. At the beginning, we notice a spike in the number of exposed people, which represents the tipping point of the system. However, we can also see that the system has a high resiliency since it quickly returns to a balanced state. In the steady state, we see an almost constant number of susceptible/exposed/infected along time; this is mainly because in the case of Malaria, people do not infect each other, but instead mosquitoes are what spread the disease. Since the contact and infection rate between people and the mosquitoes are constant, we will see this linear steady result. The width of the colour strips is directly proportional to the states' period lengths. In other words, if we increase the incubation period duration, for instance, the width of the exposed state will increase.

Parameter Set 2 Results:

In this second parameter set, we notice a progressively increasing number of deaths. We see a negative exponential change in the numbers relative to each state. This behaviour is explained by the decrease in the hospital capacity compared to the previous case. We can still see the spike in cases at the beginning of the model representing the tipping point, and we can still see how the system tries to bounce back to a stable state. However, the lack of hospital spaces is a strong disturbance, and this is why we see this exponentially increasing number of deaths. We also notice here that most of the times the hospitals will have full capacity, and compared to the previous parameter set case, the percentage of people infected at home is higher due to lack of hospital spaces (77% infected at home here versus 65% in parameter set 1). Finally, looking at the animation in the model, we can see that there is no clear pattern related to the infection; this is again because people cannot infect other people, and thus the proximity of people (being in the same household) does not lead to any significant effect on the overall system behaviour.

Parameter Set 3 Results:

Finally, in this parameter set, the contact rate and the infection probability are increased. This leads to a vigorous spike at the beginning of the model where all people get exposed rapidly. This represents the major tipping point. However, we can see in this case that the system takes a much longer time to recover until it returns to a steady state, where we notice repeated oscillations but with progressively decreasing magnitudes. We also notice an exponentially increasing number of deaths following the tipping point due to the inability of hospitals to manage the sudden and steep increase in the number of cases. This shows how the system behaves differently based on the initial strength of the tipping point. Some systems may even not have the ability to recover, and thus the effect might be irreversible.

Parameter Set 4 Results:

The behaviour of the overall system in this parameter set is very similar to what we saw in "parameter set 1", where we have a tipping point at the beginning and then a quick recovery to reach a stable state. However, the effect of the introduction of a combination of preventive policies with random percentages can easily be seen from the widths of the strips and thus the number of people in each state. We see fewer infections, which reduces the pressure on hospitals, and this most importantly leads to a significant decrease in the number of deaths, which is ultimately the most important result that is usually sought for in the context of disease control.

Section 2: Tipping Point as Complexity Concept (Marketing Model)

This model is used to study the effect of marketing campaigns such as advertisements on the sales of a product. For this particular model, the sales of medical equipment is chosen. In the model, the person will become aware of the product by a certain way, and then the person will either buy the product or will not buy it. There are three possible ways for a person to become aware of the existence of the product:

1. Hearing about the product from a friend. This friend may convey the product as good or bad. This is called "word of mouth".
2. Through watching an advertisement. This is the paid media marketing method.
3. By coincidence just by walking in front of the store or by other means of chance. This

is called the "seasonality purchase" in the model, and in this case in particular, the person will automatically buy the product.

Parameters Description:

- Contact Rate: the number of people someone might come into contact with during a day.
- Loyal Purchase Again Probability: represents the probability for someone to stay loyal to the brand, meaning that they will continue buying the product in the future.
- Time to Forget Product (in days): represents the number of days it takes to forget that the brand exists after a person becomes aware of it. This works only for people who didn't buy a product from that brand, so they forget its existence after either watching an ad or having a friend tell them about it.
- Adoption Fraction: the fraction of people who will buy the product once they become aware of it.
- Evaluation time: the time it takes for someone to either like or dislike the product after buying it.
- Paid Media Investment Fraction: this is just a fraction of effectiveness for the ad to be able to reach people to make them aware of the brand. If this fraction is increased more, people will become aware of the product.
- Fraction of people purchasing in seasonality: This is the fraction of people who will discover the product without ads or a friend telling them. They will discover it just by walking in front of the store or by other means of chance.
- Ad effectiveness: once an ad is watched by a person, this is how likely they are to buy the product.
- Product Quality: a measure that determines how likely someone is to like the product and become loyal to the brand.

- Product Lifetime: how much time it takes for someone to need the product again after purchase.

Results:

In the results side of the window, you will see two-time plots, one that shows the change in the state of the customers with time, and the other that shows the change in the amount of sales per day with time. Note that these are NOT stack charts as in the case of the SEIRS and the Malaria model. These are normal time plots that show the change in the numbers with time of each state in an independent graph.

Parameter Set 1 Results:

In this parameter set, the first thing we notice is the quick decrease in the number of unaware people of the product. This of course is the result of the advertising campaign as well as people's effect on each other through word of mouth. This point can represent a certain time after the introduction of a new product in the market, and we can see it as the tipping point that changes the overall state of the system. This, of course, is accompanied by an increase in the number of aware people, of which some decide to use the product and others decide not to. All this is reflected by the successive increase in the buyers and the never-buy graphs following the increase in the aware graph.

We notice, however, that the number of non-buyers is always bigger than buyers, and this is because, first, the contact rate is low, and second, since only 50% of the population will be reached by the ad campaign and since only a certain percentage of those reached will actually end up buying the product. We can also clearly see that the system has enough resilience to get to a stable state after the non-linearity of the tipping point; this is apparent from the relative linearity that we see in all the graphs after 150 days.

Parameter Set 2 Results:

The interesting observation in this parameter which distinguishes it from the previous one is that we don't see a decrease in the number of buyers after the beginning spike. The system stabilized at the high levels of buyers. The controlling factor here that determines this behaviour is the time to forget the product. This is a factor that affects the people who didn't buy the product, so they forget its existence after either watching an ad, or having a friend tell them about it. This time is decreased to 15 days in this parameter set, which means that when this time expires, the number of non-buyers decreases, since they are once again susceptible to advertising and to subsequent purchase of the product, and so part of these non-buyers will turn into buyers. This is why we will not see the decrease in the number of buyers since this decrease is compensated by the introduction of new buyers who were previously non-buyers. The system reaches a dynamic equilibrium following the tipping point, where internal modifications are occurring at the micro level, but we see stability at the macro level as if no change is happening because these internal modifications cancel each other out.

Parameter Set 3 Results:

In this parameter set we have a low adoption factor, and this is why a very limited increase in the number of buyers contrasted with a steep and big increase in the number of non-buyers is noticed. We also see successive oscillations with decreasing magnitudes that reflect how the system is trying to reach a stable state. These oscillations are due to the forgetting factor, where non-buyers decrease after the forgetting period since they are once again susceptible to advertising and subsequent purchase of the product. However, they increase again since the adoption factor is low, and so only a small amount of these previous non-buyers will change to buyers. The majority will remain non-buyers.

Session 3 Instructional Video Transcript

Introduction: Hello, and welcome, Medical Students. My name is Gina, and I am a lecturer in Australia who teaches in both the MD program and the Population and Global Health undergraduate and postgraduate degrees. I will be here to share my knowledge about epidemics and complex systems. Let's start learning.

Concept Definition and Description:

Two people balancing on a see-saw, hot coffee cooling down to room temperature, the same number of hawks and rabbits being born and dying in an ecosystem: these are all examples of systems at equilibrium. Systems like these tend to have a state of equilibrium or stability—people balancing, coffee at room temperature, and stable birth/death rates of hawks and rabbits—where the overall behaviour remains relatively constant even though there are dynamic interacting elements. This type of stability is referred to as *dynamic equilibrium.* It represents the system's ability to maintain an overall stability despite disturbances of the elements.

Section 1: Epidemics (COVID-19 Model)

The coronavirus, also known as SARS-CoV-2 or COVID-19, is a highly contagious virus that primarily spreads through respiratory droplets. When an infected person coughs, sneezes, talks or even breathes, these tiny droplets containing the virus can be released into the air. A person can become infected by inhaling these droplets or by touching surfaces or objects contaminated with the virus and then touching their face, particularly the mouth, nose, or eyes. Close contact with an infected person, typically within about 2 metres, poses a higher risk of transmission. It is important to note that asymptomatic individuals can also transmit the virus, making it important to practice preventative measures such as wearing masks, maintaining physical distance, frequently washing hands, and avoiding crowded indoor spaces.

Let's start with the general parameters which are similar to the SEIRS model:

- Initial Prevalence, which is the % of people that are infected at the beginning of the simulation.
- Infectivity, which is how likely they are to be infected when they are contacted.
- Fraction of people who are asymptomatic.
- Contagious Period, or how much time the person will remain contagious (infectious).
- Incubation Duration is how much time the person is exposed to the virus before becoming contagious.
- Immunity Duration is the time during which the person is immune from getting the disease.

Another section of the model addresses the sociability parameters of the COVID model, which deals with how much contact people have at work and in the recreation space.

The next section is healthcare parameters, which are mainly related to hospital capacity. This includes “Admission”, which means the number of available places for normal admission in the hospital. The “ICU” is the number of available intensive care units, and “Ventilators” refers to the number of available ventilators. People who are in need of hospitalization of any sort and cannot find available locations at the hospital are more likely to die at home.

Finally, when you tick the policies section, it displays some preventive policies that can help reduce the spread of the virus; for instance, the fraction of people using a mask and the fraction following hygienic guidelines. You can also generate a government plan, like having a quarantine, so that people are not allowed to leave their house and have to work from home. In that case you can specify the fraction of people who abide by the isolation. The second government plan is vaccination, where you can specify the fraction of people who are vaccinated.

Results:

You have three graphs in the results section:

1. A time plot showing the change of people states over time.
2. A stack chart of the people states (refer to SEIRS model for the detailed description of a state chart)
3. A time plot showing change in the number of people in each unit in the hospital over time.

Parameter Set 1 Results:

In this parameter set, we can see the typical behaviour of disease spread where we have peaks in infections followed by low infection periods. We also notice that the intensity of the peaks is gradually decreasing. These are also accompanied with similar patterns in the graph showing the number of hospital admissions, also with consecutive peaks and lows. These repetitive consecutive wave-like structures we see in the graphs show that the system is in a stable or equilibrium. We also notice an increasing number of deaths in this case. This is mainly because most infected people (70%) are symptomatic, which means that their sickness level is more severe and will need hospitalization. Because hospitals have a limited capacity, we will see a rising numbers of deaths.

Parameter Set 2 Results:

Asymptomatic persons will continue their life as usual since they are unaware of their sickness and thus the disease will spread more widely, we can see a lower rate of admissions to the hospital because the cases will be less severe. Consequently, we will see fewer deaths.

Parameter Set 3 Results:

In this final parameter set, we see a completely different behaviour. The controlling factor here is the incubation period length, which is greater here. We notice here that the peaks in the graphs are much less severe, which means that the disease is spreading at a slower rate and that the number of infected people will be almost constant over time.

Parameter Set 4 Results (Activating the policies parameters here):

In this set, the same parameters used in "parameter set 1" are repeated, but certain preventative policies are added: the use of masks, following of hygienic guidelines, and vaccinations. We see less fluctuations in this case, which means a slower rate of spread of the disease as well as fewer deaths. The disease spreads at an almost steady rate similar to the behaviour we saw in "parameter set 3", but in this case with a much lower number of infections.

Parameter Set 5 Results (Activating the policies parameters here):

This set shows the importance of quarantine as a preventative measure even when used alone. The use of quarantine helps in limiting the spread of the disease in a short period of time even when the infectivity rate in high.

Section 2: Dynamic Equilibrium as Complexity Concept (Wolf-Sheep Model)

Predator-prey models are models used to study the dynamics between two or more interacting species in an ecosystem: predators and their prey. These models provide insights into the complex relationship between predator and prey populations and how they can influence each other's population sizes over time. Predator-prey models often exhibit cyclic behaviour, where the populations of predators and prey rise and fall in a predictable pattern. Initially, an increase in prey population leads to a subsequent increase in the predator population as they have more food available. However, as the predator population grows, it puts more pressure on the prey population, causing their numbers to decline. Eventually, with fewer prey available, the predator population also declines. This cycle continues in a repeating pattern representing an equilibrium in the system. Predator-prey models, thus, can exhibit stable equilibrium points, where the populations of predators and prey remain relatively constant over time. These equilibrium points represent a balance between predator and prey numbers, where the rate of prey reproduction matches the rate of predation. However, small perturbations in the system can lead to cyclical or chaotic behaviour.

In this model, specifically, we have two sets pf predator-preys. The wolves (predator) and the sheep (prey) as well as the sheep (predator) and the grass (prey). These two sets will feed on each other, and at the same time the wolves and the sheep can reproduce, and the grass can grow again. The concept of energy is also introduced into the model. Both wolves and sheep gain certain amounts of energy when they consume their preys. This energy is what helps them to live and be able to move around. If they reach a point where they have no more energy, they will die.

Parameters Description:

- Initial sheep population: number of sheep at the beginning of the model
- Initial wolves population: number of wolves at the beginning of the model
- Initial green grass %: percentage of squares in the layout that have grass (the layout is divided into squares which can either have grass or no grass). Note: if the "use grass" option is not chosen, this is equivalent to having 100% initial green grass.
- Show energy: If this option is chosen, the energy relevant to each animal will appear in the animation (above each animal)
- Sheep reproduce probability: the probability for each of the sheep to reproduce. This is evaluated each time step.
- Wolf reproduce probability: the probability for each of the wolves to reproduce. This is evaluated each time step.

- Gain from food for sheep of wolves: You can specify the numbers here for the amount of energy each animal will gain when they consume a prey.
- Grass regrow time: the number of time steps needed for a grass to grow back in a square after it is eaten.

Results:

You will see two graphs. The first shows the change in the number of animals over time, and the second shows the change in the animal's energy over time.

Parameter Set 1 Results:

The model with this parameter set shows the perfect example of how things are balanced in nature. Entities in a system depend on each other for survival but can reach an equilibrium state where they can co-exist in harmony without one of them eliminating the other. We can see that from the results, where we have peaks in the number of each species that are consecutive; meaning that the relationship between the number of wolves and sheep is inversely proportional. In other words, the greater the number of wolves, the less the number of sheep. And the same with the sheep and the grass. This state of equilibrium can continue unless a disturbance affects the system balance.

Parameter Set 2 Results:

This set of parameters shows a disturbance in the system that leads to an eventual imbalance in resources. Here we can see that the number of sheep is increased while keeping the same % of grass as in the previous case. The sheep in this case are not able to survive due to the limited amount of resources. We can see from the graph that after a certain time, all the sheep disappear. Two factors contribute to this result: first the limited resources for the sheep to survive and, second, the fact that the wolves are consuming the available sheep.

Parameter Set 3 Results:

Here we see another set of parameters that also leads to breaking the equilibrium state. Here we start with a big number of sheep, but we also have enough grass to sustain them. The reproductive probability of sheep is high while that of wolves is low. This leads to a drastic increase in the number of sheep, as we can see from the rapidly increasing exponential graph. Although the wolves are able to survive due to the abundance of the food supply, the number of wolves stays almost constant due to the low reproductive rate. However, the overwhelming number of sheep, as it is clear from the animation, leads to lack of space and crowding, which has the effect of throwing the system out of equilibrium.

Appendix F: Pretest and Post-test

Pretest Background questions

1. What is your gender? (Male / Female)
2. How old are you?
3. What year are you in?
4. Is English the primary language spoken at home? (Yes / No)
5. Have you studied or been exposed to epidemiology before? (Yes / No)
6. What is your major?

Pretest Part A: Knowledge of Epidemics

This section asks you questions about epidemics. Please answer each question by writing a short answer and a brief paragraph explaining your response Even if you do not feel you are certain about the 'right' answer, please provide a response.

1. What are the modes of transmission of infectious organism (virus, parasite, or bacteria)? Why is it important to know them?
2. What does mortality refer to in epidemics?
3. What is the difference between latency and incubation period?
4. Please describe what disease prevalence is.
5. What are examples of dynamic equilibrium in epidemics? Please explain.
6. What are examples of tipping points in epidemics? Please explain.
7. What are examples of emergent properties in epidemics? Please explain.

Pretest Part B: Problem Solving

In this section, we give you a challenge problem to answer. However, even if you do not feel you "know" the right answer, please do your best to think of a plausible response.

Problem 1:

Department of Prevention and Infection Control in the Ministry of Health (MOH) reported that about 10,000 patients presented with common cold during the period of December to March in 2019. At the same time, Centres for Disease Control and Prevention (CDC) reported approximately 32 million ambulatory care visits with the common cold. Moreover, MOH reported that during the first two weeks of January 2019, there were about 15 cases of Tuberculosis in AL Seeb Region. Several new cases are seen in a particular area during a relatively brief period. Based on the above scenario, please answer the following question:

1.1 Please write a short essay to answer this question. What are the differences between an epidemic and an outbreak of a disease in relation to the different population sectors involved in the proliferation of the disease?

Problem 2:

Of the 40 cases of Salmonellosis caused by Salmonella Enterica in 1 week, they were traced to a single meal served at a cafeteria. This disease is considered a very common cause of gastroenteritis in the developed world, and an invasive disease in the developing world. Diagnosis relies on isolation of the organism from stool cultures or by detection of pathogen-specific nucleic acid. Please answer the following questions:

2.1. Discuss the disease and its ability to move through the population.

2.2. With the information that you have regarding this disease, what preventative measures can be put into place?

Post-test Part A: Knowledge About Epidemics

This section asks you questions about epidemics. Please answer each question by writing a short answer and a brief paragraph explaining your response. Even if you do not feel you are certain about the 'right' answer, please provide a response.

1. What are the modes of transmission of infectious organism (virus, parasite, or bacteria)? Why is it important to know them?
2. Why does mortality refer to in epidemics?
3. What is the difference between latency and incubation period?
4. Please describe what disease prevalence is.
5. What are examples of dynamic equilibrium in epidemics? Please explain.
6. What are examples of tipping points in epidemics? Please explain.
7. What are examples of emergent properties in epidemics? Please explain.

Post-test Part B: Problem Solving

In this section, we give you a challenge problem to answer. However, even if you do not feel you "know" the right answer, please do your best to think of a plausible response.

Problem 1:

Department of Prevention and Infection Control in the Ministry of Health (MOH) reported that about 10,000 patients presented with common cold during the period of December to March in 2019. At the same time, Centres for Disease Control and Prevention (CDC) reported approximately 32 million ambulatory care visits with the common cold. Moreover, MOH reported that during the first two weeks of January 2019, there were about 15 cases of Tuberculosis in AL Seeb Region. Several new cases are seen in a particular area during a relatively brief period. Based on the above scenario, please answer the following question:

1.1 Please write a short essay to answer this question. What are the differences between an epidemic and an outbreak of a disease in relation to the different population sectors involved in the proliferation of the disease?

Problem 2:

Of the 40 cases of Salmonellosis caused by Salmonella Enterica in 1 week, they were traced to a single meal served at a cafeteria. This disease is considered a very common cause of gastroenteritis in the developed world, and an invasive disease in the developing world. Diagnosis relies on isolation of the organism from stool cultures or by detection of pathogen-specific nucleic acid. Please answer the following questions:

2.1. Discuss the disease and its ability to move through the population.

2.2 With the information that you have regarding this disease, what preventative measures can be put into place?

Problem 3:

According to the 2021 statistics, Oman's 5.16 million own approximately 5.31 million mobile phones. Road injury and mortality is associated with inattention and mobile phones contribute to a large percentage of these injuries and deaths. The rate of injuries and deaths relate to the government regulations and individual behaviour.

3.1. Can you explain the relationship between mobile phones, road injury, and mortality? How does prevention relate to data surveillance?

3.2. Please describe and explain why different preventative and regulatory measures are used for different members of the population to reduce the rate of injury and mortality associated with mobile phones?

Appendix G: Self-report Measures

We are interested in your ***personal opinion*** *on the way you learned epidemics diseases during the study and the way you learn usually. There are no right or wrong answers.*

Think about how your learning experience of epidemics and complex systems concepts went in the last three sessions. Provide your feedback on what you think works well for you and what parts could be improved.

What works well (if anything): ____

What could be improved (if anything): ____

Other comments: ____

Think about how your usual learning experience of epidemics or any other theoretical medical courses normally go. Provide your feedback on what you think works well for you and what parts could be improved.

What works well (if anything): ____

What could be improved (if anything): ____

Other comments: ____

____

Appendix H: Interview Questions

Hello participants. Thank you for volunteering to participate in this focus group interview. My name is Raeda Al Hinai. As you have completed all the study materials in Moodle platform, I would like to hear your perceptions toward such learning experience by discussing some questions with the group members.

I will be recording this discussion. This interview is completely anonymous, and your responses will be confidential. Please answer the questions to the best of your ability. The duration of this interview will be approximately 45 minutes.

1. In this study, you used agent-based models (ABMs) to learn about epidemics. What did you like most and like least about learning epidemics through ABMs?
2. Contrast and compare how you learned about epidemics ideas in this study to a more traditional science teaching method?
3. Did you find the use of systems ideas to be helpful or not helpful in learning the epidemics concepts? Please explain.
4. Do you feel the approach used in this study might be helpful to you in learning other medical and health knowledge?
5. Do you have any other comments about the learning activities used in this study?

Appendix I: Marking Rubric for Pretest and Post-test

Q1: What are the modes of transmission of infectious organisms (viruses, parasites, or bacteria)? Why is it important to know them?

Scientifically correct answer:

The modes of transmission are how the virus or parasite can be transmitted from one person to another. There are many ways in which this transmission can occur, such as through direct contact (touch), in the air, by liquid droplets, through contaminated objects, etc. It is essential to understand the virus's transmission mode to limit the spread of the disease through the appropriate prevention methods. For example, if we know that a virus is airborne, masks can be used as a prevention method against the spread of the virus. There are several modes of transmission of infectious organisms:

- Direct contact, which involves touching an infected person or their bodily fluids, is one transmission mode. Additionally, some pathogens can travel through the air, posing a risk to individuals in close proximity.
- Indirect Contact Transmission: This involves a vehicle or vector-like contaminated objects (doorknobs, utensils), insects (mosquitoes, ticks), or food and water. Regular hand hygiene and disinfection practices become critical in mitigating this mode of transmission
- Airborne transmission implies that the virus or parasite can be suspended in tiny particles, making respiratory precautions such as wearing masks a valuable preventive measure.

- Droplets expelled through actions like coughing or sneezing represent another mode of transmission. These droplets may contain infectious agents and can be inhaled by others nearby. Understanding this mode prompts the recommendation for maintaining safe distances and adopting respiratory hygiene practices to reduce the risk of exposure.
- Vector-Borne Transmission: Diseases transmitted by vectors like mosquitoes, ticks, and fleas. For example, malaria is spread by mosquitoes.

By comprehending the specific mode of transmission, tailored prevention methods can be implemented. For instance, if a virus is known to spread through direct contact, promoting handwashing and avoiding close contact become essential. If it is airborne, the emphasis shifts to using masks and ensuring adequate ventilation in enclosed spaces.

In detail, the importance of Knowing Modes of Transmission can include:

- Effective Prevention and Control: Understanding how diseases spread is crucial for developing effective prevention and control strategies.
- Public Health Policies: It informs public health policies, including vaccination, sanitation, and quarantine measures.
- Healthcare Protocols: Helps establish proper infection control protocols in healthcare settings, like using PPE.
- Educational Campaigns: Aids in designing targeted educational campaigns to inform the public about how to reduce the risk of infection.
- Research and Development: Guides research and development of new diagnostic methods, treatments, and preventive measures.

Table I1

Scoring Rubric for Q1

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer	Answer left blank or smiley or question mark or	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying the same thing as in question	Viruses, parasites and bacteria are capable of transmitting infectious organisms based on their type.	Will be obvious there is no answer in the words given.
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating one idea/term correctly but having these terms explained incorrectly Use terms like susceptible, exposed, infectious No further explanation on how those terms are related	contact skin. droplets in the air. eaten in contaminated food. insects. pathogens can spread by air, contact and surfaces. it's important to know the modes of transmission for each pathogen to establish guidelines when dealing with it in personal, hospital and community levels Direct contact, as droplets in the air, contaminated in food and water, contaminated animals' stool. It's important to know all the modes of transmission in order to prevent them as much as possible.	Make sure answers stay on track. The mode of transmission doesn't change with prevention types however (personal, hospital, comm levels).
2	A correct answer with some/no scientifically correct terminology used/ correct with explanations	Use of two or more terms that are explained correctly Use terms like spread in population, model Scientifically correct explanation of how those terms are related	There are numbers of modes of transmission of pathogens such as: injection-by contaminated food, inhalation-by air droplets, direct contact and indirect contact. I think knowing the mode of transmission is the first step to eliminate the infection	This is a correct answer it just lacks depth.

Score	Answer Type	Description	Examples	Comments
			The modes of transmission can be understood for effective infection control and prevention of diseases. How an infectious agent spreads can help with reducing the risk. Things like contact, droplets, and food can all be involved.	
3	A correct answer with scientifically correct (Ph.D. level) terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) Use of scientifically correct terminology like transmission, direct and indirect contact Scientifically correct explanation of how those terms are related Presence of more terms which explain further the concept, e.g. virus, parasite, contamination, and hygiene	The modes of transmission are how the virus or parasite can be transmitted from one person to another. There are many ways in which this transmission can occur, such as through direct contact (touch), in the air, by liquid droplets, through contaminated objects, etc. It is essential to understand the virus's transmission mode to limit the spread of the disease through the appropriate prevention methods. For example, if we know that a virus is airborne, masks can be used as a prevention method against the spread of the virus. Direct contact, which involves touching an infected person or their bodily fluids, is one transmission mode. Additionally, some pathogens can travel through the air, posing a risk to individuals in close proximity. Indirect Contact Transmission: This involves a vehicle or vector-like contaminated objects (doorknobs, utensils), insects (mosquitoes, ticks), or food and water. Regular hand hygiene and disinfection practices become critical in mitigating this mode of transmission.	They must provide a good description however I do not think they need to provide every mode of transmission but at least 2 for full marks.

Score	Answer Type	Description	Examples	Comments
			Airborne transmission implies that the virus or parasite can be suspended in tiny particles, making respiratory precautions such as wearing masks a valuable preventive measure. Droplets expelled through actions like coughing or sneezing represent another mode of transmission. These droplets may contain infectious agents and can be inhaled by others nearby. Understanding this mode prompts the recommendation for maintaining safe distances and adopting respiratory hygiene practices to reduce the risk of exposure. Vector-Borne Transmission: Diseases transmitted by vectors like mosquitoes, ticks, and fleas. For example, malaria is spread by mosquitoes.	

Q2: What does mortality refer to in epidemics?

Scientifically correct answer:

Mortality is a crucial metric in epidemics because it represents the death rate or percentage within a population affected by a specific disease. In scientific terms, mortality refers to the number of deaths caused by a particular illness relative to the total population. Understanding and monitoring mortality rates are important for several reasons:

- *Severity Indicator:* Mortality rates serve as a key indicator of the severity of an epidemic. By tracking the number of deaths attributed to a specific disease, health practitioners and researchers can assess the impact on the affected population. High mortality rates may suggest a more severe and potentially dangerous outbreak, requiring urgent and targeted intervention strategies.
- *Resource Allocation:* Understanding mortality rates aids in prioritizing and allocating healthcare resources effectively, especially in crises. In detail, high mortality rates may necessitate the deployment of more medical personnel, increased hospital capacity, and the development of targeted public health campaigns. Additionally, understanding mortality helps in assessing the economic and social impact of an epidemic, informing policies that address both public health and broader societal concerns.
- *Public Health Strategies:* Mortality data inform public health strategies, including vaccination campaigns, quarantine measures, and public awareness efforts.
- *Epidemiological Tracking:* Tracking mortality rates over time helps in understanding the progression of the epidemic and the effectiveness of control measures. A rising mortality rate may indicate a worsening situation, prompting the need for additional resources, treatment options, or revised preventive measures.

- *Research and Development:* High mortality rates can drive research and development efforts towards new treatments, vaccines, and diagnostic methods. For instance, identifying the most vulnerable populations and understanding the severity of the disease aids in prioritising vaccine distribution. It also helps evaluate the efficacy of vaccines in reducing mortality rates, which is a critical factor in determining their overall success in controlling the spread of the disease.

Table I2

Scoring Rubric for Q2

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer	They are important because...	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea/ shows lack of seriousness in ideas presented /saying the same thing as in question: The text that does not answer the question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas	-Because the death usually happens among small group of people compared to the total infected people. -Because it has lots of health complications	
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but	-to measure or predict surveillance from mortality or morbidity from disease.	

Score	Answer Type	Description	Examples	Comments
		having these terms explained incorrectly Answer seems correct but does not elaborate on why she thinks that's the case	-mortality in epidemics is important as it tells us about the severity and urgency of a certain	
2	A correct answer with some/no scientifically correct terminology used / correct with explanations	Use of one or more terms that are explained correctly	It provides a measure of death in relation to a specific disease	It really needs to say in comparison to the total population so that there is a comparison group.
3	A correct answer with scientifically accurate (Ph.D. level) terminology, elaboration, and explanations.	Complete expert explanation (which shows understanding and further relationships to relevant ideas)	Mortality is a crucial metric in epidemics because it represents the death rate or percentage within a population affected by a specific disease. In scientific terms, mortality refers to the number of deaths caused by a particular illness relative to the total population. Understanding and monitoring mortality rates are important for several reasons: Severity Indicator: Mortality rates serve as a key indicator of the severity of an epidemic. By tracking the number of deaths attributed to a specific disease, health practitioners and researchers can assess the impact on the affected population. High mortality rates may suggest a more severe and potentially dangerous outbreak, requiring urgent and targeted intervention strategies.	The student does not need to provide all of the reasons however an example of a few helps demonstrate their understanding.

Score	Answer Type	Description	Examples	Comments
			Resource Allocation: Understanding mortality rates aids in prioritizing and allocating healthcare resources effectively, especially in crises. In detail, high mortality rates may necessitate the deployment of more medical personnel, increased hospital capacity, and the development of targeted public health campaigns. Additionally, understanding mortality helps in assessing the economic and social impact of an epidemic, informing policies that address both public health and broader societal concerns. Public Health Strategies: Mortality data inform public health strategies, including vaccination campaigns, quarantine measures, and public awareness efforts. Epidemiological Tracking: Tracking mortality rates over time helps in understanding the progression of the epidemic and the effectiveness of control measures. A rising mortality rate may indicate a worsening situation, prompting the need for additional resources, treatment options, or revised preventive measures. Research and Development: High mortality rates can drive research and development efforts towards new treatments, vaccines, and diagnostic methods. For instance, identifying the most vulnerable populations and understanding the severity of the disease aids in prioritising vaccine distribution. It also helps evaluate the efficacy of vaccines in reducing mortality rates, which is a critical factor in determining their overall success in controlling the spread of the disease.	

Q3: What is the difference between latency and incubation period?

Scientifically correct answer:

The distinction between latency and the incubation period is essential in understanding the progression of infectious diseases within the host's body.

Incubation Period:

- Definition: The incubation period is between exposure to an infectious agent and the onset of symptoms. It represents the time it takes for the pathogen to replicate and reach a sufficient level within the host, leading to the onset of clinical symptoms.
- Significance: It is critical for understanding the disease progression timeline and for effective epidemiological tracking.
- Public Health Impact: Helps determine the appropriate duration for quarantine or isolation and guides policies on tracing and testing after exposure.

Latency Period:

- Definition: The latency period refers to the time between exposure to the infectious agent and when the person becomes contagious, i.e., capable of spreading the disease to others. The virus remains dormant or inactive during this latent phase, not causing apparent symptoms. This period can be characterised by the virus residing in a latent state, potentially integrating into the host's genetic material or persisting in a non-replicating form.
- Latency is often associated with certain types of viruses, such as herpesviruses or retroviruses, which can establish long-term infections in the host. In these cases, the virus can remain latent for extended periods, only becoming active and replicating under specific conditions or triggers.

- Significance: It is essential for understanding and controlling the spread of the disease.
- Public Health Impact: As latency is a dynamic aspect of viral infections, implementing strategies like contact tracing and social distancing is crucial to prevent transmission from asymptomatic carriers.

Key Differences:

- The incubation period is about the onset of symptoms, whereas the latency period is about the capacity to transmit the disease.
- A person can be infectious during the latency period even if they are asymptomatic.
- Control measures may differ based on whether the focus is on preventing symptom onset (incubation) or transmission (latency).

Table I3

Scoring Rubric for Q3

Score	Answer type	Description	Examples	Comments
NA	No answer/ Off task	No answer	"Don't know"	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea/shows lack of seriousness in ideas presented/saying the same thing as in question	They are similar stages but one comes before the first	
			How a disease is tracked through computer networks.	
	Off-topic	• The text that does not answer the question or part of it.		

Score	Answer type	Description	Examples	Comments
1	Partly correct or incomplete /no scientifically correct terminology used / correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly • The answer seems correct but does not elaborate on why • Ideas presented are correct but appear to lack correct understanding	Latency: organism is hiding in the body and not replicating. Incubation Period: organism is hiding and replicating.	Answers that are not quite correct or somewhat unknown.
2	A correct answer with some/no scientifically correct terminology used / correct with explanations	Use of one or more terms that are explained correctly • Use of relevant • Elaborate on the fit into the field	latency period individual does not suffer from any illness incubation period: infectious period. Latency: the infectious organism does not replicate Incubation: where the infectious organism replicates	Short but brief, almost shows understanding but not enough depth to truly decipher this.
3	A correct answer with scientifically correct terminology, elaboration, and explanations.	Reasonably full expert explanation (which shows understanding and further relationships to relevant ideas) • Explains the definition and role in the field	incubation period refers to the period from the start of the infection to the time in which symptoms of the disease starts appearing. while latency period is a property of certain pathogens in which it can stay in the human body for a long time without causing a disease, but a disease can be triggered in immunosuppressed or immunocompromised patients causing their symptoms to come back ending the latent	I don't think the students have to cover all aspects here but show a true understanding of their link.

Score	Answer type	Description	Examples	Comments
		• Includes definition and shows how it fits into a multidisciplinary context	Latency refers to the time between the initial infection and manifestation of symptoms. Some infections, like HIV, can have a long latency period when the virus replicates within the body without any noticeable symptoms. Whereas the incubation period is when the individual becomes symptomatic after being exposed to a disease-causing agent. In brief, the time between being exposed to the flu and the onset of flu symptoms.	

Q4: Please describe what disease prevalence is.

Scientifically correct answer:

Definition:

Disease prevalence is a fundamental epidemiological concept that provides insight into the burden of a particular disease within a given population. It is expressed as the percentage or ratio of individuals in that population with a specific disease at a particular time. In brief, it measures how widespread a disease is in a specific population.

Calculation:

Prevalence is calculated by dividing the number of existing cases of a disease by the total population at risk. This metric offers a snapshot of the overall frequency of the disease within the defined population. There are two main types of prevalence: point prevalence, which measures the proportion of individuals with the disease at a specific point in time, and period prevalence, which considers the occurrence of the disease over a defined period.

Importance:

- High disease prevalence suggests a substantial presence of the disease in the population, indicating a greater public health impact and potential strain on healthcare resources. This information is vital for policymakers, as it influences decisions related to healthcare infrastructure, the allocation of funds, and the development of targeted prevention and intervention strategies.
- Low disease prevalence does not necessarily diminish the significance of a disease, especially if it is severe or has long-term consequences. In such cases, even a low prevalence may significantly impact affected individuals and the healthcare system.

- Prevalence is a dynamic measure that can be influenced by various factors, including changes in the incidence (rate of new cases), treatment efficacy, and population demographics. It is often used in conjunction with other epidemiological measures, such as incidence rates and mortality rates, to provide a comprehensive understanding of the disease's impact on a population.

Table I4

Scoring Rubric for Q4

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer	I do not know.	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea/shows lack of seriousness in ideas presented /saying the same thing as in question • The text that does not answer the question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	It's the rate of disease. It's how you measure health. A measure of how likely a person is to get sick.	Random draws at showing an understanding.
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly • Ideas presented are correct but appear to lack correct understanding	-Frequency of the disease in a certain community or group -it is how common is a disease in a certain area	

Score	Answer Type	Description	Examples	Comments
2	A correct answer with some/no scientifically correct terminology used /correct with explanations	Use of relevant terms relating two ideas/terms correctly • Ideas presented are scientifically correct	it is the number of people having the disease to the total number of people	
3	A correct answer with scientifically correct terminology, elaboration, and explanations.	Complete expert explanation (which shows understanding and further relationships to relevant ideas)	Ratio of individuals who suffer from disease to all community people including those people at a point in time	I really think showing that at a point in time or particular time is key to full marks. It is not over time or without a parameter.

Q5: What are examples of dynamic equilibrium in epidemics? Please explain.

Scientifically correct answer:

Dynamic equilibrium in epidemics refers to a state where the rates of new infections and the rates of recovery (or removal, due to either recovery or death) are equal, leading to a stable number of active cases over time. This concept is crucial in understanding how diseases spread and are controlled within populations. Here are some examples:

- Endemic Diseases: Diseases that are constantly present in a population, like the common cold, often reach a dynamic equilibrium. The number of new infections is roughly equal to the number of people recovering or developing immunity, so the total number of infected individuals remains relatively stable.
- Seasonal Fluctuations: Certain diseases like influenza show seasonal patterns. They might reach a dynamic equilibrium during peak seasons, where new infections balance out recoveries and then decline as the season ends.
- Controlled Epidemics with Vaccination: When a significant portion of a population is vaccinated, an epidemic can reach a dynamic equilibrium. New cases continue to occur but at a rate balanced by the number of people recovering or being protected through vaccination.
- Disease Eradication Efforts: In efforts to eradicate diseases like polio or measles, dynamic equilibrium can be an intermediate stage. Here, the disease is not spreading rapidly but is maintained at a low level due to ongoing transmission in unvaccinated or vulnerable populations.
- Herd Immunity: This occurs when a large portion of a population becomes immune to an infectious disease, either through previous infections or vaccination, reducing the

spread. This can lead to a dynamic equilibrium where the disease persists at a low level, with new infections balanced by recoveries.

- HIV/AIDS Management: With effective antiretroviral therapy, HIV has reached a dynamic equilibrium in many populations. The rate of new infections is balanced by the rate of people effectively managing the disease, leading to a stable number of people living with HIV.
- Animal Reservoirs and Zoonotic Diseases: Diseases that jump from animals to humans can reach a dynamic equilibrium in their animal hosts, maintaining a constant reservoir of the pathogen.

In each of these cases, the dynamic equilibrium is influenced by factors like the rate of transmission, the effectiveness of public health interventions, changes in population immunity, and environmental factors. Understanding and managing these equilibriums are key in controlling infectious diseases.

Table I5

Scoring Rubric for Q5

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer		
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	Vaccinations and requirements needed to avoid or live with the infection. Epidemics aims to create a disease-free world Result of many components interacting such as cells interact to make tissue, and tissues interact to make organ and organs to organism	Answers have irrelevant content, or they are off task.
1	Partly correct or incomplete /no scientifically correct terminology used / correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly • Shows limited understanding of certain concepts •	New infections and recovery are the same	Here, there is an indirect indication of time. It also has a limited understanding of concepts the main concepts.
2	A correct answer with some/no scientifically	Use of one or more terms that are explained correctly, like rate, infection and recovery. It	It refers to a state when rate of new infections and rates of recovery are equal – so a stable number of people in the population have the disease	Lacks elaboration or an example to demonstrate an understanding

Score	Answer Type	Description	Examples	Comments
	correct terminology used / correct with explanations	will include what it is without much elaboration or further links.		
3	A correct answer with scientifically correct terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology. Use terms like rate, infection, and transmission. • Includes at least one example to elaborate concept • Possible further link • Some abstraction of ideas	• A balance between factors that promote the spread of a disease and those that work against it. Like during the flu season, more people are inside because of the weather which increase the rate of transmission but measure like vaccinations, hand washing helps decrease the spread of the virus.	Using an example to demonstrate the understanding of the key 'balance'

Q6: What are examples of tipping points in epidemics? Please explain.

Scientifically correct answer:

Tipping points in epidemics signify critical thresholds or changes in the levels of certain parameters within the system. These transitions often occur in response to a small change or the cumulative impact of previous incremental changes. Understanding and identifying these tipping points are essential for devising effective public health strategies and interventions. Here are some examples and explanations:

- Threshold of Herd Immunity: This is a fundamental tipping point in epidemics. Herd immunity occurs when a sufficient proportion of the population becomes immune to an infectious disease, either through vaccination or previous infections, thus providing indirect protection to those who are not immune. When the immunity level crosses a certain threshold, it significantly reduces the spread of the disease, potentially leading to its decline. For example, in the case of measles, the herd immunity threshold is high, requiring about 95% immunity to prevent outbreaks.
- Behavioural Changes: Shifts in public behaviour can act as a tipping point. For example, the adoption of widespread mask-wearing, social distancing, and hand hygiene can dramatically reduce transmission rates. Conversely, relaxing these measures too quickly can lead to a resurgence of the disease.
- Mutation of the Pathogen: The emergence of a more transmissible or virulent strain can be a tipping point. This was observed with the COVID-19 pandemic, where variants like Delta and Omicron changed the dynamics of the epidemic, leading to new waves of infections.
- Super-Spreader Events: Certain events or settings can facilitate the rapid spread of an infectious disease, acting as tipping points. For instance, a large gathering where

preventive measures are not followed can lead to a significant outbreak, amplifying the spread of the disease.

- Resource Thresholds: In healthcare settings, reaching capacity limits (e.g., ICU beds, ventilators, healthcare personnel) can be a tipping point. Once these resources are overwhelmed, the mortality rate may increase, not just from the epidemic disease but also from other health conditions that cannot be adequately treated.
- Public Health Interventions: The implementation or withdrawal of public health measures (like lockdowns and travel restrictions) can serve as tipping points. The timing and effectiveness of these interventions can drastically alter the epidemic curve, so identifying these tipping points is important for public health planning, as it helps gauge the impact of implemented measures and informs decisions on resource allocation and the timing of intervention strategies.

Table I6

Scoring Rubric for Q6

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer		
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	Diseases need to be identified in different ways to be better researched and understood. The more we know about diseases, the more we can do things to stop others from getting it. The two types of diseases are important because viruses can be divided into either or ruled out if they are not contagious and won't spread.	Answers have irrelevant content, or they are off task.
1	Partly correct or incomplete /no scientifically correct terminology used / correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly • Shows limited understanding of certain concepts	Left-sided – when more people are catching the disease Right-sided – when more people become immune	It also has a limited understanding of concepts the main concepts.
2	A correct answer with some/no scientifically correct terminology used / correct with explanations	Use of one or more terms that are explained correctly • includes what it is without much elaboration or further links	Threshold of herd immunity Super-spreader event small things happen during the epidemic and impact potentially in the spread of the disease like the number of initial infected	Lacks elaboration or touches on some of the right information but not quite.

Score	Answer Type	Description	Examples	Comments
			people and the mode of transmission of the microorganism, maybe the population in that area	
3	A correct answer with scientifically correct (Ph.D. level) terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology. Use terms like transmission, proportion, threshold and behavioural change. • Includes at least one example to elaborate concept • Possible further link • Some abstraction of ideas	• A critical threshold or change within the system. For example, herd immunity occurs when a large enough proportion of the population has become immune. As more individuals become immune, the transmission of the disease slows down in the population. This helps protect even those who are not immune.	Key is that it is a critical threshold or change in the parameters of the system. Answer must include an example and understanding in relation to disease transmission.

Q7: What are examples of emergent properties in epidemics? Please explain.

Scientifically correct answer:

In epidemiology, emergent properties refer to complex and often unexpected phenomena that arise from interactions within a population during an epidemic. These properties go beyond the simple sum of individual components and result from the collective behaviour of the elements involved. Emergent properties help researchers and public health professionals understand the dynamic and often nonlinear nature of infectious disease spread within populations. Several emergent properties can be observed:

- Outbreak Patterns: Patterns like clustering of cases or waves of infection emerge from the interactions between individuals, their mobility, and social behaviours.
- Herd Immunity: This occurs when a significant portion of a population becomes immune to a disease, reducing its spread and providing indirect protection to the entire population. This collective immunity emerges from individual immune responses.
- Resistance Patterns: The emergence of antibiotic or antiviral resistance is a complex interplay of pathogen evolution, medication usage, and healthcare practices.
- Social Behaviours: Changes in social behaviour in response to an epidemic, like increased handwashing or vaccine hesitancy, are emergent properties resulting from individual and collective perceptions and decisions.
- Epidemic Thresholds: The point at which an outbreak becomes an epidemic or a pandemic is an emergent property, depending on factors like infection rate, population density, and travel patterns.

- Health System Response: The overall response of healthcare systems, including resource allocation and policy changes, emerges from the interactions of various components of the healthcare infrastructure.
- Evolution of Virus Strains: The emergence of new virus strains with different properties, such as increased transmissibility or resistance to vaccines, is an emergent property driven by the evolutionary dynamics of the virus.

Table I7

Scoring Rubric for Q7

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer		
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	The dynamics of how a disease is spread A disease moves quickly from the vector to the host. The speed at which this occurs is referred to as dynamic.	
1	Partly correct or incomplete /no scientifically correct terminology used / correct without explanations	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly • Use of relevant terms such as vector and host. • The answer seems correct but does not elaborate • Ideas presented are correct but appear to only include part of the answer	This includes understanding how a population interacts during an epidemic. This relates to not just the person but the group.	• Uncertainty about the answer • Uncertainty about the term used

Score	Answer Type	Description	Examples	Comments
2	A correct answer with some/no scientifically correct terminology used / correct with explanations	Use of one or more terms that are explained correctly • Use of relevant terms such as interactions, vector and host.	This relates to the causes of interactions within a population during an epidemic that are erratic and not fully explainable.	Touching on the full concept but not providing examples of the types.
3	A correct answer with scientifically correct terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology • Use of relevant terms such as interactions, vector and host. • Talk about emergent properties from a higher level, not just the specifics.	These are the characteristics of behaviours that come from a collective level of a population and are usually not easily predicted or explained by individual components alone. This can include things like outbreak patterns, herd immunity, resistance patterns, and social behaviours to name a few. These would help identify the dynamic nature of infectious diseases.	Make sure they name at least 3 to get full marks and explain them as a concept.

Problem-solving Questions

1. **Please write a short essay to answer this question. What are the differences between an epidemic and an outbreak of a disease in relation to the different population sectors involved in the proliferation of the disease?**

Scientifically correct answer:

To understand the differences between an epidemic and an outbreak in the context of disease proliferation across different population sectors, it's essential first to define these terms and then examine their implications in varied demographic and geographic contexts.

An outbreak refers to the occurrence of more cases of a disease than expected in a given area or among a specific group of people over a particular period of time. Outbreaks can occur in a limited geographic area or can extend over several countries. They can also be limited to a specific demographic, such as children, healthcare workers, or residents of a nursing home. An outbreak does not necessarily have a large number of cases; it can involve a small number or even a single case in a new or particularly sensitive environment (e.g., a rare disease appearing in a new region).

On the other hand, an epidemic is defined as an increase, often sudden, in the number of cases of a disease above what is normally expected in that population in that area. The key difference from an outbreak is the scale of the spread. An epidemic covers a wider geographic area, affecting a large number of people. It often involves multiple population sectors and can cross socio-economic boundaries, impacting diverse groups such as urban and rural populations, wealthy and poor, or various age groups.

The proliferation of disease in different population sectors during an outbreak or epidemic can be influenced by several factors:

- Demographic Susceptibility: Certain demographics may be more susceptible to a disease due to factors like age, pre-existing health conditions, or genetic predispositions. For example, an outbreak of a respiratory illness might disproportionately affect the elderly or those with compromised immune systems.
- Socioeconomic Factors: Socioeconomic status can influence exposure risk, access to healthcare, and the ability to engage in disease prevention measures. In an epidemic, lower-income areas might experience higher transmission rates due to crowded living conditions, limited access to healthcare, or a lack of resources for proper sanitation.
- Geographic Distribution: The spread of a disease can vary significantly between urban and rural areas. Urban areas, with higher population density and greater mobility, might see a rapid spread of disease, while rural areas might experience delayed but concentrated outbreaks due to limited healthcare infrastructure.
- Behavioural Practices: Cultural, occupational, and social practices can influence disease spread. For instance, a disease might spread rapidly in a sector where social gatherings are common or among healthcare workers due to their exposure to infected patients.

In summary, while an outbreak might be confined to a specific group or location, an epidemic spread more broadly, impacting multiple population sectors. Understanding these distinctions is crucial for public health officials in devising targeted strategies for disease surveillance, containment, and treatment. The scenario presented, with common cold cases and Tuberculosis in specific areas and time frames, can be analysed through these lenses to understand the dynamics of disease spread in different population sectors.

Table I8

Scoring Rubric for Problem-solving Q1

Score	Answer Type	Description	Example
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it.	The difference between an epidemic and an outbreak lies with the type of disease that occurs and can be tracked back to either a disease that cannot be spread or not.
1	Partly correct or incomplete /no scientifically correct terminology used	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly	An outbreak is like a small wave, and an epidemic is like a big tsunami. An outbreak happens when a few people in a small area get sick, like a small wave that comes and goes. On the other hand, an epidemic occurs when many people in a big area get sick, like a massive tsunami affecting many people and areas.

2	A correct answer with some/no scientifically correct terminology used	Use of one or more terms that are explained correctly
		• Idea looks clear and related to;
		  • Infectious disease
		  • Distinguish the difference
		• Answer is logically correct but no use of scientifically correct terminology
		Includes a weak example
			An outbreak of a disease typically affects a small, localised population, while an epidemic involves the spread of an infectious disease across a larger geographic area and population sectors. Epidemics can cause more health issues and affect the well-being of entire populations and may require more extensive measures to contain them.
			An example of an outbreak would be when many people in the same place get sick from the same thing around the same time. For example, a few students in a school may get sick with the flu, but if many more students and teachers start getting sick, too, it can become an outbreak. The school might need to close for a while to prevent the sickness from spreading even more.

Score	Answer Type	Description	Example
			An example of an epidemic is when many people in a larger area or even a whole country get sick with the same disease. For instance, if a new virus spreads quickly through a city and many people start getting sick, and then it spreads to other cities and countries, it can become an epidemic. Doctors and scientists may need to work together to find a cure or vaccine to stop the epidemic from getting worse and spreading even more.
3	A correct answer with scientifically correct terminology and explanation	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology • Give a correct explanation of the concept related to epidemics and outbreaks Include at least one correct example of	Epidemics and outbreaks are two related terms used to describe the spread of infectious diseases, but they differ in their scale and impact on population sectors. An outbreak of a disease is typically localized to a specific geographic area, such as a single neighbourhood or city, and affects a relatively small population sector. In contrast, an epidemic is a more widespread outbreak that affects multiple regions, countries, or even continents and can impact a broader range of population sectors. The critical difference between an outbreak and an epidemic lies in the size and scope of the affected population sectors. Outbreaks typically occur in small, localized populations, such as schools, hospitals, or nursing homes, and can be contained relatively quickly through measures such as quarantine, contact tracing, and targeted vaccination. Epidemics, on the other hand, involve the spread of disease across large populations, including communities, cities, and even entire countries. As such, they often require more extensive measures, such as mass vaccination campaigns, public health education, and the mobilization of emergency response resources. Another important factor distinguishing epidemics from outbreaks is the severity level and impact on different population sectors. Epidemics can have a significant impact on the health and well-being of entire populations, with the potential for high morbidity and mortality rates. In addition to the direct health consequences,

Score	Answer Type	Description	Example
			epidemics can also have significant economic and social impacts, disrupting healthcare systems, supply chains, and the functioning of entire communities. Outbreaks, while still potentially serious, tend to be more localized and contained, with a smaller overall impact on population sectors beyond the immediate affected area. An example of an outbreak is E. coli, which occurred in the United States in 2018. The outbreak was linked to romaine lettuce grown in Yuma, Arizona, and affected 210 people in 36 states. Although the number of cases was relatively small compared to other outbreaks, the outbreak received significant media coverage due to its potential to cause severe illness and the widespread distribution of the affected product. The outbreak also prompted a recall of the lettuce and highlighted the importance of food safety measures to prevent the spread of foodborne illness. An example of an epidemic is COVID-19, which travelled globally, spreading quickly and killing millions.

2a. Discuss the disease and its ability to move through the population.

Scientifically correct answer:

Salmonellosis, caused by the bacteria Salmonella enterica, is a significant public health concern due to its ability to spread rapidly through populations. To discuss this disease and its transmissibility, we need to consider several key aspects:

1. Transmission Mechanisms:
   - Foodborne Transmission: The primary mode of transmission is through the consumption of contaminated food. Common sources include undercooked meat, eggs, and dairy products, as well as fruits and vegetables that have been contaminated. In the given scenario, the disease's spread through a meal at a cafeteria is a classic example of foodborne transmission.
   - Person-to-Person: While less common, it can also spread from person to person, especially in settings with poor hygiene practices. This is particularly concerning in crowded environments like schools, nursing homes, or hospitals.
2. Infectious Dose and Incubation Period:
   - Salmonella enterica has a relatively low infectious dose, meaning that ingesting even a small amount of the bacteria can cause illness.
   - The incubation period ranges from several hours to two days, allowing for rapid onset of symptoms after exposure.
3. Symptoms and Severity:
   - The symptoms of Salmonellosis typically include diarrhea, fever, and abdominal cramps. Severe cases can lead to dehydration, which can be particularly dangerous in vulnerable populations like the elderly, infants, and those with weakened immune systems.

- In developing countries, invasive disease forms, such as typhoid fever, can occur, leading to more severe health complications.

4. Public Health Implications:

- Rapid Spread: Given the nature of its transmission and the communal settings where food is often consumed (like cafeterias), there's a high potential for rapid spread among a population.
- Outbreak Investigation: Identifying the source of an outbreak, as in the case of the cafeteria, is crucial for controlling the spread. Public health officials conduct epidemiological investigations to trace the source and implement control measures.

5. Control and Prevention:

- Hygiene Practices: Regular hand washing and proper food handling are key in preventing the spread of Salmonellosis.
- Food Safety Regulations: Ensuring that food is cooked to the proper temperatures and preventing cross-contamination in food preparation areas are essential measures.
- Public Awareness: Educating the public about the risks and prevention methods can significantly reduce the incidence of Salmonellosis.

6. Diagnosis and Treatment:

- As mentioned, diagnosis is typically through stool cultures or nucleic acid detection. Rapid and accurate diagnosis is crucial for effective treatment and for containing the spread.
- Treatment primarily involves hydration and electrolyte balance. In severe cases, antibiotics may be required.

In summary, the ability of Salmonella enterica to move through a population is facilitated by its transmission via contaminated food and the potential for person-to-person spread in certain conditions. Its rapid onset and the severity of symptoms, especially in vulnerable populations, highlight the importance of stringent food safety practices and public health surveillance to contain and prevent outbreaks.

Table I9

Scoring Rubric for Problem-solving Q2a

Score	Answer Type	Description	Example
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	Salmonellosis is a disease caught by eating salmon that is infected with a disease. This can move through the population as the people purchasing fish usually get the fish from the same batch or location, and they are probably infected by the same disease.
1	Partly correct or incomplete /no scientifically correct terminology used	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly	Salmonellosis is an infection that can be spread through germs of people. Some people are more inclined to get it as they have had it earlier, making them more susceptible to future cases.
2	A correct answer with some/no scientifically	Use of one or more terms that are explained correctly • Idea looks clear and related to; • Salmonellosis	Salmonellosis is an infection that causes mild symptoms but can also lead to serious life-threatening illness. An example of how the disease can move through the population is if someone with the infection does not wash their hands properly after using the toilet, they can

Score	Answer Type	Description	Example
	correct terminology used	• Present the symptoms or cause • Answer is logically correct but no use of scientifically correct terminology Includes a weak example	spread the disease to other people through contact with surfaces, such as door handles or handrails.
3	A correct answer with scientifically correct terminology and explanation	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology • Give a correct explanation of salmonellosis Include at least one correct example or show a statistical understanding of the disease.	Salmonellosis is a bacterial infection caused by Salmonella, which can cause a wide range of symptoms, from mild diarrhea to severe and life-threatening illnesses. The disease is highly infectious and can be transmitted through contaminated food, water, or contact with infected animals. While the disease can affect anyone, some groups of the population are more susceptible to the infection, including young children, the elderly, and people with weakened immune systems. Salmonellosis is a common foodborne illness in developed countries and a significant public health concern. According to the Australian Government Department of Health, in 2019, 4,193 cases of salmonellosis were reported, while in 2020, 3,433 cases were reported. The number of deaths due to salmonellosis in Australia is relatively low, with an average of around ten deaths reported each year. Outbreaks of salmonellosis have been linked to a wide range of food products, including eggs, meat, dairy products, and fruits and vegetables. Salmonellosis has the ability to move through the population due to its highly infectious nature and the widespread distribution of contaminated food products. The bacteria can survive in various environments, including raw and undercooked meat, poultry, and eggs, as well as unpasteurized dairy products. Contaminated food products can be distributed to

Score	Answer Type	Description	Example
			various locations, including supermarkets, restaurants, and institutions, which can increase the risk of infection to a larger population. In addition, it can spread rapidly in settings such as hospitals, long-term care facilities, and schools, where people are in close contact with each other. Poor hygiene and sanitation practices in these settings can contribute to the spread of the disease, highlighting the importance of appropriate infection control measures to prevent and control outbreaks. Salmonellosis is a highly infectious bacterial infection that can move through the population, affecting people of all ages and backgrounds. The widespread distribution of contaminated food products and poor hygiene and sanitation practices in certain settings can contribute to the spread of the disease. Effective prevention and control measures, such as proper food handling, hygiene practices, and appropriate infection control measures, are crucial to mitigate the impact of salmonellosis on the population.

2b. With your information regarding this disease, what preventative measures can be put into place?

Scientifically correct answer:

Preventive measures for Salmonellosis, particularly in the context of foodborne transmission like the scenario involving a cafeteria, involve a multi-faceted approach that includes public health strategies, food safety practices, and individual behaviours. Here are some key preventative measures:

1. Food Handling and Preparation:
   - Cooking Food Thoroughly: Salmonella bacteria are destroyed at high temperatures, so cooking foods, especially meats, eggs, and poultry, to the recommended temperatures is crucial.
   - Avoiding Cross-Contamination: Using separate cutting boards and utensils for raw meat and other foods and ensuring that surfaces and utensils are thoroughly cleaned after use can prevent the spread of bacteria.
   - Proper Storage of Food: Refrigerating or freezing perishable food promptly and defrosting foods safely is important in preventing bacterial growth.
2. Hygiene Practices:
   - Hand Washing: Regular and thorough hand washing, especially after handling raw food, using the bathroom, or before eating, is a critical preventive measure.
   - Sanitation: Keeping kitchen areas clean and regularly disinfecting surfaces where food is prepared can reduce the risk of contamination.
3. Public Health Surveillance and Response:
   - Regular Inspections: Regular health inspections of food service establishments can ensure compliance with food safety regulations.

- Rapid Response to Outbreaks: Quick identification and response to outbreaks, including tracing the source and implementing quarantine measures if necessary, are essential to prevent further spread.

4. Education and Awareness:
    - Public Education Campaigns: Raising awareness about the importance of food safety, symptoms of Salmonellosis, and when to seek medical help can empower individuals to take preventive actions.
    - Training for Food Handlers: Providing training about safe practices for those involved in food preparation and handling can significantly reduce the risk of foodborne illnesses.
5. Vulnerable Populations:
    - Extra Precautions for High-Risk Groups: People with weakened immune systems, the elderly, children, and pregnant women should take extra precautions, like avoiding raw or undercooked eggs and meats.
6. Regulatory Measures:
    - Food Safety Regulations: Implementing and enforcing strict food safety regulations and standards for food production, processing, and distribution.
7. International Standards and Cooperation:
    - Adhering to International Standards: Following guidelines set by international bodies such as the World Health Organization (WHO) and the Food and Agriculture Organization (FAO) for food safety.

Combining these individual, community, and regulatory approaches can significantly minimize the risk of Salmonellosis outbreaks. It’s also essential for individuals to be aware of the symptoms and to seek medical attention if they suspect they have been infected, especially if symptoms are severe.

Table I10

Scoring Rubric for Problem-solving Q2b

Score	Answer Type	Description	Example
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it. Part of the sentence may be correct but incorrectly related to the later part, which shows a lack of conceptual understanding/misunderstanding of the ideas.	Given salmonellosis has to do with eating salmon, just refrain from eating fish. Perhaps take all fish out of the diet to ensure you do not get it from another fish.
1	Partly correct or incomplete /no scientifically correct terminology used	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly	To prevent salmonellosis, wash hands frequently with soap and water, cook food thoroughly, avoid cross-contamination, and store food at the appropriate temperature.

2	A correct answer with some/no scientifically correct terminology used	Use of one or more terms that are explained correctly
		• Idea looks clear and related to;
		  • Brief prevention
		  • Present the symptoms or cause
		• Answer is logically correct but no use of scientifically correct terminology
			It is essential to follow good hygiene and food safety practices to prevent salmonellosis. These include washing hands before handling food, cooking food thoroughly, keeping raw and cooked foods separate, and washing hands and surfaces after handling raw meat, poultry, or eggs. It is also important to regularly clean and disinfect surfaces and utensils that come into contact with food and to store food at the appropriate temperature. These simple

Score	Answer Type	Description	Example
			steps can help prevent the spread of the bacteria that cause salmonellosis and reduce the risk of infection.
3	A correct answer with scientifically correct terminology and explanation	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology • Give more than one option for prevention Include at least one correct example or show a statistical understanding of the disease.	Particularly related to food safety and personal hygiene. One effective approach is to ensure that food is cooked thoroughly, particularly meat, poultry, and eggs, which are commonly associated with salmonellosis outbreaks. Adequate cooking kills the bacteria, reducing the risk of infection. Additionally, it is important to avoid cross-contamination by keeping raw and cooked foods separate and washing hands and surfaces thoroughly after handling raw meat, poultry, or eggs. As well, personal hygiene practices can also help prevent the spread of salmonellosis. These practices include washing hands thoroughly with soap and warm water before handling food, after using the bathroom, and after contact with animals or animal products. Regular handwashing is essential in settings such as hospitals, long-term care facilities, and schools, where people are in close contact with each other. Proper sanitation practices can also play a critical role in preventing salmonellosis outbreaks. This includes regularly cleaning and disinfecting surfaces and utensils that come into contact with food and ensuring that food is stored at the appropriate temperature to prevent the growth of bacteria. A multifaceted approach to salmonellosis prevention is necessary, including appropriate food handling, hygiene practices, and proper sanitation measures. Public education and awareness campaigns can also effectively promote these preventative measures, helping to reduce the incidence of salmonellosis in the population.

3a. Can you explain the relationship between mobile phones and road injury and mortality and how prevention relates to data surveillance?

Scientifically correct answer:

The relationship between mobile phone usage and road injury and mortality is a significant public health concern. This relationship is multifaceted, involving behavioural, technological, and policy-related aspects. Prevention strategies often hinge on understanding these dynamics through data surveillance.

1. Relationship between Mobile Phones and Road Safety:
    - Distraction: The primary issue with mobile phone use while driving is distraction. Engaging with a phone—whether it's texting, calling, browsing, or even using navigation apps—diverts attention from the road.
    - Reaction Time: Using a mobile phone while driving can significantly increase reaction time, akin to the impairment observed with intoxication.
    - Visual Attention: Mobile phone use affects drivers' visual attention, often leading them to miss critical cues, signals, and changes in traffic patterns.
    - Cognitive Load: Even hands-free mobile phone use increases cognitive load, reducing the mental resources available for safe driving.
2. Data Surveillance and Its Role in Prevention:
    - Accident Analysis: By collecting data on road accidents, researchers can identify the prevalence of mobile phone use as a contributing factor. This data can be gathered from police reports, insurance claims, and even direct observation studies.

- Trend Identification: Data surveillance helps identify trends and patterns in road accidents, including the times of day, types of roads, and driver demographics most associated with mobile-phone-related accidents.
- Policy Formulation: With robust data, governments, and road safety authorities can develop targeted policies, such as stricter laws against mobile phone use while driving, increased penalties, or awareness campaigns.
- Effectiveness of Interventions: Post-implementation surveillance helps assess the effectiveness of these policies and interventions, allowing for adjustments and improvements.

3. Prevention Strategies:

- Legislation and Enforcement: Implementing and strictly enforcing laws that prohibit the use of mobile phones while driving.
- Awareness Campaigns: Educating the public about the dangers of mobile phone use while driving through various media and public awareness campaigns.
- Technology Solutions: Promoting the use of technologies that restrict mobile phone use while driving, such as drive mode apps or in-vehicle systems that limit phone functionality.
- Cultural Shift: Encouraging a societal change in attitudes towards mobile phone use while driving, similar to how attitudes towards drunk driving have evolved.

4. Role of Individual Behaviour:

- Self-Regulation: Encouraging drivers to self-regulate their mobile phone usage, recognizing the risks involved.
- Promoting Responsible Behaviour: Through educational programs, especially targeting young and new drivers, to instill safe driving habits from the start.

In summary, the relationship between mobile phone usage and road injury and mortality is significant, with mobile phones being a primary source of driver distraction leading to accidents. Data surveillance plays a crucial role in understanding this relationship and informing effective prevention strategies. By combining legislative, educational, technological, and cultural approaches, the negative impact of mobile phone use on road safety can be mitigated.

Table I11

Scoring Rubric for Problem-solving Q3a

Score	Answer Type	Description	Example
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it.	If you have enough money to own a car, you will also own a mobile phone. In order to work and maintain a family life, you will probably use the mobile while driving, so you have more chances of getting in a car accident and having an injury. Prevention includes watching how many people own mobiles and encouraging these people not to drive.
1	Partly correct or incomplete /no scientifically correct terminology used	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly	Mobile phones can cause death when used while driving as they can be distracting for drivers. If we monitor through the available data, we may be able to determine the number of accidents and try to stop them.

2	A correct answer with some/no scientifically correct terminology used	Use of one or more terms that are explained correctly
		• Idea looks clear and related to;
		  • Brief prevention
		  • Present the basics for data surveillance
		• Answer is logically correct but no use of scientifically correct terminology
			Mobile phones have been found to be a major contributor to road injury and mortality because they can distract drivers and pedestrians from paying attention to their surroundings. Using a mobile phone while driving or walking can cause a lack of focus and can lead to accidents.
			Prevention of mobile phone-related road injuries and mortality involves data surveillance, which includes monitoring and analysing data related to accidents caused by the use of mobile phones. This data can be used to develop ways to decrease accidents with the community and legal solutions.

3	A correct answer with scientifically correct	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas)	Road injury and mortality represent a significant public health concern worldwide, with mobile phone use being increasingly recognized as a contributing factor. The use of mobile

Score	Answer Type	Description	Example
	terminology and explanation	• Use of scientifically correct terminology • Give more than one option for prevention Include at least one correct example of how data surveillance can be employed.	phones while driving, particularly for texting or browsing the internet, has been shown to increase the risk of road accidents, injuries, and fatalities. Prevention of mobile phone-related road injuries and mortality can be achieved through a combination of education, policy, and surveillance measures. Public awareness campaigns can be implemented to promote safe driving practices, including the dangers of using mobile phones while behind the wheel. Policymakers can also introduce legislation and regulations that restrict the use of mobile phones while driving and enforce penalties for violations of these laws. Data surveillance can be essential in preventing mobile phone-related road injuries and mortality by providing information on driver behaviour and crash patterns. This information can be used to develop targeted interventions and evaluate the effectiveness of prevention strategies. Data surveillance can be achieved through various methods, including traffic cameras, smartphone applications, and sensors integrated into vehicles. While data surveillance can potentially improve road safety, there are also concerns regarding privacy and data protection. Therefore, any surveillance measures must be implemented in a manner that respects individual privacy rights and adheres to ethical and legal standards. Prevention of mobile phone-related road injuries and mortality requires a multifaceted approach that includes education, policy, and surveillance measures. Data surveillance can be a valuable tool in preventing such incidents, but it must be implemented in a responsible and ethical manner. Ultimately, these prevention efforts aim to reduce the incidence of mobile phone-related road accidents and promote safer driving practices for all road users.

3b. Please describe and explain why different preventative and regulatory measures are used for different members of the population to reduce the rate of injury and mortality associated with mobile phones.

Scientifically correct answer:

Different preventative and regulatory measures are targeted at various segments of the population in addressing the issue of mobile phone-related road injuries and mortality due to the diverse risk profiles, behaviours, and needs of these groups. Tailoring these measures allows for more effective risk mitigation and behaviour modification. Here are some reasons and examples of these differentiated approaches:

1. Age-Related Differences:
    - Young Drivers: Younger drivers, often more tech-savvy and more likely to engage in risky behaviours, might require more aggressive awareness campaigns, educational programs in schools, and strict enforcement of mobile phone usage laws. They might be more responsive to digital media campaigns or technology-based solutions.
    - Older Drivers: For older drivers, the focus might be more on educational initiatives that highlight the risks of mobile phone use while driving, given that they might be less aware of the potential dangers or less adept at multitasking.

2. Experience and Skill Level:
    - New Drivers: Novice drivers might have less experience and skill handling distractions. Therefore, specific regulations like a complete ban on mobile phone use (including hands-free) for this group can be effective.
    - Professional Drivers: For commercial or professional drivers who spend more time on the road, employers can implement policies that restrict mobile phone use during driving, along with regular safety training.

3. Cultural and Social Factors:

- Community-Based Interventions: In communities where mobile phone use while driving is culturally normalized, community-led interventions and local advocacy can be more effective.
- Social Influence Campaigns: Targeting opinion leaders or influencers within communities to set examples of safe driving practices.

4. Technological Accessibility and Use Patterns:

- Tech-Heavy Users: For populations heavily reliant on smartphones, interventions might include technological solutions like apps that restrict phone usage while driving.
- Low Tech-Use Populations: For those less dependent on technology, traditional methods like billboards or radio announcements might be more effective.

5. Economic and Occupational Considerations:

- Employment Policies: Employers can play a role by establishing strict no-phone policies for employees while driving, especially in occupations that involve frequent vehicle use.
- Insurance Incentives: Offering lower insurance premiums for drivers who use phone-blocking technology or have a record of safe, non-distracted driving.

6. Legal and Regulatory Frameworks:

- Laws Tailored to Risk Levels: Laws can be more stringent for high-risk groups, like higher fines for young or new drivers caught using phones.
- Enforcement Strategies: Different enforcement strategies, such as random checks or the use of technology to detect phone use while driving, can be more effective for certain segments of the population.

7. Accessibility of Public Education and Awareness Campaigns:

- Language and Communication Strategies: Using multiple languages or communication strategies to reach diverse populations within a country.

By recognizing and addressing these differences, preventative and regulatory measures can be more effectively designed and implemented to reduce the rate of injury and mortality associated with mobile phone use while driving. Tailoring these measures acknowledges the unique characteristics and needs of different segments of the population, leading to more successful outcomes in road safety initiatives.

Table I12

Scoring Rubric for Problem-solving Q3b

Score	Answer Type	Description	Example
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying same thing as in question • The text that does not answer question or part of it.	
1	Partly correct or incomplete /no scientifically correct terminology used	Use of relevant terms without explanation or further elaboration OR relating two ideas/terms correctly but having these terms explained incorrectly	

2	A correct answer with some/no scientifically correct terminology used	Use of one or more terms that are explained correctly
		• Idea looks clear and related to;
		  • Brief reasoning behind different measures
		  • Present the symptoms or cause
		• Answer is logically correct but no use of scientifically correct terminology
			Different preventative and regulatory measures are used for different members of the population to reduce the rate of injury and mortality associated with mobile phones because different groups have different levels of risk and exposure to mobile phone-related accidents.
			
			For example, younger people may be more likely to use their mobile phones while walking or driving and may benefit from learning that emphasizes the dangers of distracted walking or driving. In contrast, older adults may be more distracted while driving and may benefit from stricter enforcement of laws that prohibit the use of mobile phones while driving.

Score	Answer Type	Description	Example
3	A correct answer with scientifically correct terminology and explanation	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology • Give more than one option for prevention Include at least one correct example or show a statistical understanding of the disease.	Regulatory measures must be tailored to different members of the population based on their unique levels of risk and exposure to mobile phone-related accidents. Various demographic groups, such as younger individuals, may be more likely to use mobile phones while walking or driving, making them particularly susceptible to accidents caused by distraction. In such cases, educational campaigns can effectively raise awareness of the dangers of distracted walking or driving. In contrast, older adults may be more vulnerable to distraction while driving and may require stricter enforcement of laws prohibiting mobile phone use while driving to prevent accidents. Specific industries like transportation or manufacturing may necessitate specific preventative and regulatory measures to reduce mobile phone-related accidents. These measures can include workplace policies that prohibit mobile phone use while operating machinery or driving and technology-based solutions like hands-free devices that enable communication without compromising visual or manual attention. Reducing mobile phone-related accidents requires a nuanced understanding of the specific risk factors and characteristics of different population groups. By developing tailored interventions that address the unique needs of different populations, we can effectively prevent accidents caused by mobile phone use and promote safety across diverse communities.

Appendix J: Thematic Analysis of Challenge Problems 1-3 Qualitative Data

Table J1

Theme 1: Number and Variety of Ideas Generated

Session	Challenge Problem	Sub-Themes	Description	PF (Experimental) Group Quote	DI (Control) Group Quote
Session 1: Epidemic & Forest Fire Models	Challenge Problem 1 (SEIRS Model – Epidemic Spread)	Biological & Epidemiological Factors	Students discussed how immunity, reinfection, and latent periods affect disease spread.	*"The disease spreads when immunity is not permanent. Infected who recover can get sick again."*	*"Some people don't present any symptoms of the disease, so they transmit it to others."*
		Human Behaviour & Social Factors	Social behaviours like handshaking, mask-wearing, and quarantine were identified as factors influencing transmission.	*"In some cultures, shaking hands is common. if they know it spreads disease, they reduce this habit."*	*"People who do not follow social distancing and hygiene rules make the disease spread faster."*
		Mathematical & Simulation-Based Insights	Students referred to how different variables in the model affected infection curves.	*"If the number of people contacted per day is high, the disease stays longer in the population."*	*"When infection probability increases, the number of cases grows rapidly."*
	Challenge Problem 2 (Forest Fire Model – Emergent Behaviour)	Fire Spread Analogies	Some students compared fire spread to disease transmission.	*"The fire here follows the same concept as disease spread. The disease does not move from one person to another, but spreads from the host, giving many new sources/hosts."*	*"The spread of fire depends on how close the trees are, just like the spread of disease depends on population density."*
		Mathematical & Simulation-Based Insights	Students attempted to explain how fire spreads in the simulation.	*"The fire itself is not moving from place to another but spreading in different directions."*	*"Fire looks like spreading because new trees catch fire, but nothing is actually moving."*

		Environmental & External Influences	Some students recognised external factors like wind and humidity affecting fire spread.	*"The direction of the winds affects the fire movement."*	*"Dry weather increases fire spread just like low immunity increases disease spread."*
Session 2: Malaria & Marketing Models	Challenge Problem 1 (Malaria Model – Disease Transmission)	Disease Transmission Mechanisms	Students identified vectors, incubation periods, and reinfection risks.	*"Malaria spreads through the bite of infected female mosquitoes."*	*"If a mosquito bites infected person, it can transmit the parasite to a healthy person."*
		Environmental & Population Factors	Factors such as population density and stagnant water were discussed.	*"If there are a lot of mosquitoes in one area, the number of infected people will increase."*	*"Mosquitoes breed in water. So, if more water and lakes, it means more mosquitoes and spread disease."*
		Prevention & Control Strategies	Students mentioned mosquito nets, repellents, and vaccines.	*"Using nets and insect sprays can prevent people from getting malaria."*	*"Malaria cases drop when people use mosquito repellents and protective clothing."*
	Challenge Problem 2 (Marketing Model – Consumer Behaviour)	Adoption & Market Penetration	Students discussed how consumers adopt new products.	*"People are more likely to buy a product if they see others using it."*	*"Advertisement and social influence help products spread faster."*
		Psychological & Social Influences	The role of advertisement and peer influence was highlighted.	*"A product becomes popular when enough people talk about it or recommend it."*	*"Marketing campaigns create demand by making a product look important."*
		Tipping Points & Product Longevity	Students explored how sales trends change over time.	*"At a certain point, sales reach peak, and then decline."*	*"Products lose popularity when new alternatives enter the market."*
Session 3: COVID-19 & Wolf-Sheep Models	Challenge Problem 1 (COVID-19 Model – Disease Spread)	Direct & Indirect Transmission Pathways	Students identified airborne spread and surface contamination.	*"COVID-19 spreads through respiratory or sneezing and contact with infected places."*	*"A person can be infected without symptoms and without knowing, they spread the disease."*

	Public Health Policies & Behavioural Changes	The role of mask mandates, lockdowns, and vaccinations was discussed.	*"Social distancing helped reduce infections during the pandemic."*	*"Quarantine and isolation helped control COVID-19 spread."*
	Long-Term Immunity & Reinfection Cycles	Students debated the impact of immunity duration.	*"Some people lose immunity after some time and can get reinfected."*	*"Vaccines provide temporary protection, but booster shots are needed to avoid for long."*
Challenge Problem 2 (Wolf-Sheep Model – Population Dynamics)	Predator-Prey Interactions	Students explored how wolves hunt sheep and maintain balance.	*"Wolves eat sheep, and when there are too many wolves, they run out of food."*	*"If wolves eat too many sheep, they will finally starve."*
	Ecological Equilibrium & Resource Availability	The balance between species populations was discussed.	*"When the number of sheep increases, wolves have more food, but if there are too many wolves, the sheep population reduces."*	*"A stable ecosystem requires a balance between predators and prey."*
	Impact of External Factors	Climate and resource availability were considered.	*"If there is not enough grass, sheep numbers drop, affecting the whole ecosystem."*	*"Droughts or winters can reduce sheep and wolf populations."*

Table J2

Theme 2: Conceptual Struggles

Session	Sub-Themes	Description	PF (Experimental) Group Quote	DI (Control) Group Quote
Session 1: Epidemic & Forest Fire Models	Conceptual Struggles (Immunity, reinfection misconceptions)	Some students misunderstood immunity and reinfection cycles.	*"The disease spreads in waves because people get infected, recover, then lose immunity, and get infected again forever."*	*"Immunity does not stop the disease, as even immune people can get sick again immediately."*

	Perceptual & Model Interpretation Struggles (Visual illusions in fire model)	Some students struggled to interpret the fire spread mechanism in the simulation.	*"The fire spreads because it jumps between trees, just like a person moving around spreads disease."*	*"The trees don't move, but the fire does, like an infection traveling through the air."*
	Reasoning & Explanation Struggles (Overgeneralizations, lack of causal connections)	Some students made vague or inaccurate connections between fire and disease models.	*"Fire spreads the same way disease does; both move from one place to another."*	*"The more trees there are, the faster fire spreads, like how a crowded area spreads disease faster."*
Session 2: Malaria & Marketing Models	Conceptual Struggles (Misunderstanding of vector-borne transmission)	Some students struggled to understand that malaria requires a mosquito vector.	*"Malaria spreads from person to person just like COVID-19."*	*"People can catch malaria directly if they are near an infected person."*
	Perceptual & Model Interpretation Struggles (Misinterpretation of tipping points)	Some students misinterpreted tipping points in disease spread.	*"Malaria cases randomly go up and down because the disease disappears and comes back."*	*"If too many people get infected, the disease suddenly stops spreading."*
	Reasoning & Explanation Struggles (Failure to connect Socioeconomic& behavioural patterns)	Students struggled to explain how socioeconomic factors impact malaria prevention.	*"If people don't use mosquito nets, they will all get sick, no matter what."*	*"If hospitals are full, people won't go to them, and that will make the disease disappear faster."*
Session 3: COVID-19 & Wolf-Sheep Models	Conceptual Struggles (Difficulty in understanding immunity & virus evolution)	Some students misinterpreted the role of immunity in virus spread.	*"COVID-19 spreads forever because people never develop full immunity."*	*"The virus mutates immediately, so vaccines are useless."*
	Perceptual & Model Interpretation Struggles (Population balance misconceptions in Wolf-Sheep model)	Some students misunderstood how wolves and sheep interact in the simulation.	*"Wolves hunt sheep until there are none left, then they die too."*	*"If sheep keep increasing, wolves stop hunting and eat grass instead."*
	Reasoning & Explanation Struggles (Failure to explain tipping points in disease & predator-prey interactions)	Some students struggled to identify when and why population levels change suddenly.	*"The number of sheep and wolves always stays the same, no matter what."*	*"If too many sheep are eaten, they all disappear, but the wolves don't die."*

Table J3

Theme 3: Relevance of Ideas to the Problem

Session	Sub-Themes	Description	PF (Experimental) Group Quote	DI (Control) Group Quote	Key Insights
Session 1: SEIRS & Fire Models	Understanding of Core Model Mechanics	Some students demonstrated correct understanding of epidemic and fire models.	*"Fire doesn't move, but it spreads by changing tree states, like infection spreads by changing immune status."*	*"The disease spread is modeled by probabilities, just like the fire simulation."*	Both groups showed evidence of recognizing the core mechanics in the models. PF students related state transitions to infection spread, while DI students correctly referenced probability and susceptibility dynamics.
	Appropriateness of Analogies & Real-World Connections	Some students made strong but imperfect analogies between models.	*"COVID-19 spreads like fire in a dry forest, but people can stop it by using protection."*	*"Quarantine is like removing trees from the fire path so it can't spread."*	Students attempted to map concepts across domains, such as fire and disease spread. PF students used behavioural interventions as analogies for fire breaks, while DI students connected quarantine to structural changes in system spread. While creative, some analogies overlooked model-specific limitations.
Session 2: Malaria & Marketing Models	Understanding of Core Model Mechanics	Some students correctly linked malaria spread to vector behaviour.	*"Mosquitoes spread malaria, so preventing mosquito bites reduces infections."*	*"The tipping point in malaria spread is when too many mosquitoes are infected."*	DI students gave more structured responses; PF students focused on causal mechanisms.
	Appropriateness of Analogies & Real-World Connections	Some students made insightful but imperfect marketing connections.	*"A product spreads like a disease—once enough people buy it, it becomes popular."*	*"Marketing campaigns work like vaccines, stopping people from switching brands."*	PF students drew broader analogies; DI students linked marketing to public health strategies.
Session 3: COVID-19 & Wolf-Sheep Models	Understanding of Core Model Mechanics	Some students accurately described virus transmission and population balance.	*"If people lose immunity, COVID-19 can come back, just like wolves return when there are more sheep."*	*"The predator-prey balance works like a controlled epidemic cycle."*	Both groups grasped cyclical population dynamics and disease resurgence.

	Appropriateness of Analogies & Real-World Connections	Some students made strong but incomplete connections between models.	*"If too many wolves hunt, they will run out of sheep, just like if a virus spreads too fast, it burns out."*	COVID-19 waves are kind of like the wolf-sheep cycle, they go up when conditions are changing and drop when things change, creating rise and fall over time.	PF students focused on depletion analogies; DI students connected to real-world wave patterns.

Table J4

Compare-Contrast Task, Session-wise Themes

Session	Theme	Sub-Theme	Description	Control Group Quotes	Experimental Group Quotes	Key Insights
Session 1	Structural-Level Features Description: *Identifying deeper structural aspects, such as causal, dynamic, and emergent patterns within the models.*	Emergent Properties	The behaviours and patterns that arise from interactions within the system rather than from individual components.	*"They both look at emergent properties of something that spreads."*	*"Both models are examples of complex systems. They demonstrate emergences and contain important parameters that control the microlevel of the system."*	Similarity: Both models focus on emergent behaviours of spread.
		Factors Influencing Spread	The key elements that determine how and where the spread occurs, such as density, proximity, resistance, and external conditions.	*"All of them depend on the main factor that may affect the spread. Fire depends on the distance, and the SEIRS model depends on the characteristics of the disease."*	*"In both models, the SEIRS model and the forest fire model, the number of objects increases when the density and closeness to each other increase. However, in the SEIRS model, the infection prevalence rate may be affected by factors like the infectious organism."*	Difference: Fire spread is spatially dependent, while disease spread depends on biological and immunity factors.

	Model Complexity & Outcomes	Differences in structural design, components, and the range of possible behaviours and results in each model.	*"Forest fire model is a one-way process with one outcome, but the SEIRS model has multiple outcomes."*	*"In the SEIRS Model, there is more variables to test, like the incubation period."*	Difference: SEIRS allows for multiple outcomes, while the forest fire model is linear and irreversible.
	Transmission & Repeatability	How transmission occurs, whether it can repeat, and the conditions affecting reinfection	*"Once a tree is burned, it cannot be burned again, so no repeating cycle like SEIRS."*	*"The SEIRS model allows reinfection and has a latency period where no transmission occurs at this period."*	Difference: SEIRS allows reinfection and latency, while fire spread is permanent and irreversible.
	Immunity Differences	How resistance develops, whether it is permanent or temporary, and how it affects future spread.	*"Fire spread is irreversible, while SEIRS includes immunity and recovery."*	*"The more the resistance of trees or the more immunity of people, the less the spread of the disease or fire. In both models, when the density of people or trees increases, the fir or disease spreads more."*	Difference: SEIRS includes immunity and recovery; fire spread does not allow recovery.
Surface-Level Features Description: *Identifying similarities or differences that are visible and superficial without addressing the core principles of the models.*	Shared Visual Features	The visual elements used in both models, such as simulations, graphs, and interactive elements, to represent spread.	*"The spread of fire in the forest is like the spread of disease among people."*	*"Both give a visual representation of the spread. In both models, we can adjust some variables."*	Similarity: Both models use simulations and allow variable adjustments, but SEIRS has more visual data analysis (charts, ratios).
			"Both have simulation video, and in both you can control the number of contacted people (or tree density) and infection probability (tree heat resistance)."	*"In the SEIRS model, we can view the results in the form of a chart."*	
				"In the forest model, we can find the ratio."	

Session 2	Structural-Level Features Description: *Explain deeper causal, dynamic, and emergent patterns that drive tipping points in the models.*	Impact of External Conditions	How external factors (e.g., environment, policies, marketing strategies) influence tipping points in both models.	*"Following the policies of prevention of malaria will help to reduce the risk of malaria."*	*"In this parameter: Paid media investment fraction: A tipping point occur when the investment fraction rise to a critical point, and as a result increase customer engagement and sales."*	Tipping points in both models depend on external forces; public health policies (malaria) or marketing investment (product adoption).
		Role of Preventive Measures	How interventions (healthcare policies, advertising campaigns) influence tipping points in each system.	*"Tipping points can be seen in the context of hospital and ICU capacity. If the number of malaria cases exceeds the capacity of healthcare services, there is a tipping point which causes limited resources and high mortality rates."*	*"Tipping points help make smart decisions by knowing what changes in customer behaviour, sales or product quality can make big difference."*	Malaria interventions aim to reduce spread, while marketing interventions maximize product adoption; both prevent tipping points from causing collapse.
		Self-Sustaining Spread	Whether and how each system experiences reinforcement mechanisms that lead to continued spread or adoption.	*"Tipping points have appear in the models due to some changes at different macro and micro levels in which we saw a huge change after that points."*	*"Tipping points may occur when the adoption fraction of a product reaches a critical level."*	Once tipping points are crossed, both systems reinforce themselves; disease spreads or product adoption accelerates.
		Thresholds and Critical Mass	A tipping point occurs when key thresholds (infection rates, product adoption) are exceeded, leading to irreversible shifts.	*"In malaria model the tipping points are, hospitalization, ICU, and Death."*	*"In the marketing model: The tipping point is when there are a dramatic decrease of the number of people that was not aware of the product."*	Both models show threshold; effects once exceeded, spread becomes difficult to control (malaria) or adoption becomes rapid (marketing).

	Surface-Level Features Description: *Capture observable effects and measurable indicators of tipping points in the models.*	Key Drivers of Tipping Points	Specific variables (e.g., infection rates, consumer adoption rates) that determine when a tipping point is reached.	*"Tipping points can be seen if the infection period takes longer than usual and requires stronger control policies."*	*"Tipping points may occur when the adoption fraction (number of customers who buy the product) reaches a high level."*	Malaria tipping points are driven by biological factors (infection probability), while marketing tipping points rely on consumer behaviour (brand memory).
		Observable Triggers of Tipping Points	Clear signs that indicate when a system is approaching or has reached a tipping point.	*"In malaria model sometimes there will be a very increase in the number of infected people and gradually it becomes low number."*	*"In marketing sometimes there will be high demand in a certain product and after sometime it becomes unneeded."*	Both models show sudden spikes before stabilizing; infections peak before intervention, and sales peak before leveling off.
		Rapid Spread & Exponential Growth	How both models exhibit periods of exponential growth before stabilizing or shifting to a new state.	*"The peak occurred at the beginning of the disease spread. As the disease started to infect more people, the community started to control the disease until returned to a stable state."*	*"In the marketing model, sales spike when people first learn about the product. After that, sales drop and reach a stable level. New peak can happen when advertisements are showing again."*	Both models experience exponential growth phases; either in deaths (malaria) or repeat customers (marketing).
Session 3	Structural-Level Features Description: *Explain deeper causal, dynamic, and*	Control Mechanisms	How elements within each system interact, and how interventions modify system behaviour.	*"Both systems show connections within their populations."*	*"In both cases, the behavior individuals or sheep and wolves impacts the dynamics of the overall system."*	Both models exhibit interdependence—COVID-19 through human interaction, Wolf-Sheep through predator-prey balance.

emergent patterns shaping system behavior.	Model Complexity & System Dynamics	Differences in how complex interactions shape each system	*"Covid-19 model is more complex than the Wolf-Sheep model with some features. Covid-19 model has policies to control infection, but the other model does not have any policies to decrease the problem effect."*	*"The Covid-19 model had much more comprehensive parameters."*	COVID-19 involves more variables and interventions; Wolf-Sheep reflects simpler natural dynamics."
	Resource Dependency & Population Balance	How resources affect system stability and growth.	*"Both systems need to ensure that there is enough resources available to live like hospital services (COVID-19) and sheep (wolf). If those resources decrease, it leads to death."*	*"The major difference is the Wolf-Sheep model is found in an ecosystem and is difficult to change the dynamic, while the COVID-19 model faces many interventions to avoid increasing deaths."*	Wolf-Sheep relies on ecological balance, whereas COVID-19 requires policy interventions.
	Equilibrium & Stability Mechanisms	How each system maintains balance or collapses under different conditions.	*"Th main similarity between the two systems is how different factors affect equilibrium. In COVID-19, the changes of patients when recovering and dying can show the equilibrium. But in the Wolf-Sheep model, there is balance between different species."*	*"Initially, both systems involve equilibrium. In the COVID-19 model, increase in infections is followed by a decline. The Wolf-Sheep model shows an increase in prey leads to an increase in predators, then causing a decline in sheep, followed by a decline in wolves."*	Both models show dynamic equilibrium, but in COVID-19 it is shaped by human interventions, while in Wolf-Sheep it arises from natural predator-prey cycles.
Surface-Level Features Description: *Capture observable effects, visual representations, and*	Graphical & Visual Representation	How each model is displayed and interpreted through visuals.	*"Both show graphical representation."*	*"Both show graphical representation."*	Both models use graphical simulations to show system changes, with COVID-19 emphasizing infection trends and Wolf-Sheep highlighting population cycles.

measurable system changes.	Spatial & Environmental Configuration	How space and environment shape interactions within each model.	*"The Wolf-Sheep model allows movement in 2D space that makes individuals look separate, whereas the COVID-19 model work on the probability of whether people around you affected you or not."*	*"Deaths are not counted in the predator model but are shown through the rise and decline in population."*	Wolf-Sheep incorporates spatial movement, while COVID-19 is more probability-based.
	Nature of Entities & Domain Context	The fundamental difference in what the models simulate (human vs. ecological systems).	*"The COVID-19 model is a human disease model, while the Wolf-Sheep model involves animal species."*	*"The difference is that in the first model we are dealing with people's lives, while the second one is dealing with sheep and wolves (the animal cycle)."*	

Appendix K: Marking Rubric for the Session-by-Session Challenge Problem 3

Challenge Problem 3 (Q1- Session1):

You are a scientist who works in the public health department of a Middle Eastern country. The Minister of Health has asked you to advise on the dangers associated with a highly infectious disease that has recently been identified, with unknown characteristics. Provide the Minister a summary of your assessment on how the disease is spread amongst the population depending on the disease characteristics. Please present your summary in general terms so that the public would understand.

Scientifically correct answer:

We are looking at a disease that could be spreading from person to person. This could be via direct or indirect contact. The dangers associated with this form of spread is that if it is very virulent it could spread quickly through regular interactions. If it is airborne transmitted, which could be from droplets or the particles from droplets. This would mean sharing the same space as someone infected could result in transmission. This would mean it would be hard to identify who gets infected as it would take a brief interaction or passing to spread.

Table K1

Scoring Rubric for Challenge Problem Q3

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer	Answer left blank or unrelated response.	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying the same thing as in question	The disease spreads through a combination of microscopic changes that are unknown and coincidences.	Just say words that aren't related to the question.
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating one idea/term correctly but having these terms explained incorrectly • Use terms like spread, person to person, contact type, airborne. • No further explanation on how those terms are related	The disease spreads due to a mix of factors, including transmission through close contact and environmental conditions, but further details are needed to fully understand its dynamics.	Starting to kind of getting the idea but there is not enough to determine completeness.
2	A correct answer with some/no scientifically correct terminology used/ correct with explanations	Use of two or more terms that are explained correctly • Use terms like spread, person to person, contact type, airborne. • Scientifically correct explanation of a type of spread (i.e., face to face or airborne)	The disease can spread when people are in close contact with each other and more so if they are in a crowded space.	The idea is starting to evolve but still not there.

Score	Answer Type	Description	Examples	Comments
3	A correct answer with scientifically correct (Ph.D. level) terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use of scientifically correct terminology like spread, person to person, contact type, airborne. • Scientifically correct explanation of how those terms are related • Presence of more terms which explain further the concept, e.g. airborne transmission and air drops and proximity, surfaces	are looking at a disease that could be spreading from person to person. This could be via direct or indirect contact. The dangers associated with this form of spread is that if it is very virulent it could spread quickly through regular interactions. is airborne transmitted, which could be from droplets or the particles from droplets. This would mean sharing the same space as someone infected could result in transmission. This would mean it would be hard to identify who gets infected as it would take a brief interaction or passing to spread.	They must provide a clear understanding that it has to do with transmission types. They do not have to provide all the examples but be confident showing at least one.

Challenge Problem 3 (Q1- Session 2): You are a researcher working at the local university and you are asked for recommendation policies to implement in response to the Malaria situation based on your current research findings. Think about the recommendations you can make and their implication in the way the population will react compared to the results in reducing the spread of this disease. Explain your reasoning for a non-technical audience, using epidemics and complex systems concepts as to why your policy recommendation will be effective in the short and/or long term.

Scientifically correct answer:

Factors to be considered include:

- Preventative policy - What is reasonable for individuals to do. Is it reasonable to shut all the windows and doors in a house without air conditioning? Is it better to add netting over the beds to keep sleeping areas safe from mosquitos and wear clothing to cover the body and wear mosquito repellent. Can natural plants be burned to reduce mosquitos? The population will react to reasonable requests, but not expensive unreasonable ones.
- If you use prevention techniques it can then reduce the incidence of malaria and then less mosquitos become carriers which reduces malaria.
- Education of the population can also help to reduce malaria as an understanding of the mechanism of transmission and how to prevent this can increase chances of prevention occurring.
- A malaria prevention policy typically includes measures such as distributing mosquito nets treated with insecticide, indoor residual spraying of insecticides, ensuring access to anti-malarial medications, promoting community awareness about prevention

methods, and improving healthcare infrastructure for early diagnosis and treatment. These policies aim to reduce the transmission of malaria and mitigate its impact on affected populations.

Table K2

Scoring Rubric for Challenge Problem G7

Score	Answer Type	Description	Examples	Comments
NA	No answer/ Off task	No answer	Answer left blank or unrelated response.	
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying the same thing as in question	Get the government to ban mosquitos.	Lots just say words that aren't related to the question OR repeat of the question.
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating one idea/term correctly but having these terms explained incorrectly • Use terms that show either an understanding of the prevention policy	Prevention policy from the university.	Starting to get that there is a concept but the concept itself is wrong or it is not a developed answer.
2	A correct answer with some/no scientifically correct terminology used/ correct with explanations	• Use of two or more terms that are explained correctly • Use terms that show either the epi side or at least the selling side like behaviour.	Develop a policy that involves giving all people education about preventing malaria and how to do it. This could be like explaining the impact of the disease and how things like nets treated with insecticide can prevent mosquitos get in and also prevent malaria from occurring.	The idea is starting to develop a prevention policy but without a fully described understanding so just parts.

Score	Answer Type	Description	Examples	Comments
3	A correct answer with scientifically correct (Ph.D. level) terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use terms that show either the epi side or at least the selling side like behaviour, contact rate, media, seasonality, quality, etc. • Understand what prevention is.	Preventative policy - What is reasonable for individuals to do. Is it reasonable to shut all the windows and doors in a house without air conditioning? Is it better to add netting over the beds to keep sleeping areas safe from mosquitos and wear clothing to cover the body and wear mosquito repellent. Can natural plants be burned to reduce mosquitos? The population will react to reasonable requests, but not expensive unreasonable ones. ou use prevention techniques it can then reduce the incidence of malaria and then less mosquitos become carriers which reduces malaria. cation of the population can also help to reduce malaria as an understanding of the mechanism of transmission and how to prevent this can increase chances of prevention occurring. alaria prevention policy typically includes measures such as distributing mosquito nets treated with insecticide, indoor residual spraying of insecticides, ensuring access to anti-malarial medications, promoting community awareness about prevention methods, and improving healthcare infrastructure for early diagnosis and treatment. These policies aim to reduce the transmission of malaria and mitigate its impact on affected populations.	They must provide a clear understanding what a prevention policy is and give at least one example of one and how it would impact the disease.

Challenge Problem 3 (Q1- Session 3): As a Director of the Inpatient ward at the local hospital, analyse the way hospitals respond during an increase in symptomatic populations. Propose some appropriate recommendations for the healthcare professionals to reduce the mortality in the population.

Scientifically correct answer:

During an increase in symptomatic populations, hospitals must strategically allocate resources to effectively manage patient care and mitigate the spread of infection. This includes increasing staffing levels by mobilizing additional healthcare professionals, such as nurses, doctors, and support staff, to handle the influx of patients. Moreover, expanding bed capacity by repurposing non-critical care areas, setting up temporary wards, and streamlining discharge processes helps accommodate the rising number of admissions. Ensuring an ample supply of critical medical resources, including personal protective equipment (PPE), ventilators, oxygen, medications, and other essential equipment, is crucial for delivering comprehensive care.

Triage and prioritization protocols play a pivotal role in managing patient flow during surges. Hospitals implement strict triage protocols to prioritize care based on the severity of patients' conditions. Establishing rapid assessment units enables swift evaluation and stabilization of patients before admitting them to appropriate wards. Stringent isolation measures for contagious patients and enhanced cleaning and disinfection practices across all hospital areas are essential to curb infection spread. Visitor restrictions are also enforced to minimize the risk of transmission within the hospital environment.

Effective communication and coordination are vital for optimizing hospital response efforts. Internally, clear and frequent communication among hospital staff facilitates coordinated patient care and resource allocation. Externally, maintaining open communication channels with public health authorities, other healthcare facilities, and emergency services ensures collaborative efforts and resource sharing during crises.

Supporting the well-being of healthcare professionals is imperative to sustain efficient care delivery. Providing mental health resources and support helps manage stress and prevent burnout among staff. Implementing adequate rest periods and shift rotations ensures a healthy and resilient workforce capable of meeting increased demands.

To reduce mortality rates, healthcare professionals emphasize early intervention and continuous monitoring of symptomatic patients. Early detection and prompt treatment initiation are prioritized to prevent complications. Adherence to evidence-based clinical guidelines and standardized treatment protocols ensures consistency in care delivery and enhances patient outcomes. Leveraging telemedicine for remote management of less severe cases and implementing remote monitoring tools support efficient resource allocation and patient management within hospital settings.

A multidisciplinary approach to patient care involves collaborating with specialists from various fields to address complex cases comprehensively. Coordinating care among different departments and integrating patient education on symptom management, early treatment, and infection prevention measures are integral components of holistic patient support. Data-driven decision-making, facilitated by rigorous data collection and analysis, informs ongoing improvements in care processes and resource utilization. Continuous capacity building through training sessions and simulation drills prepares healthcare teams to effectively manage surges in symptomatic populations and maintain high standards of care delivery during public health emergencies.

Table K3

Scoring Rubric for Challenge Problem Q12

Score	Answer Type	Description	Examples	Comments
0	Incorrect response/frivolous answer/ irrelevant ideas/cyclic answer	Scientifically irrelevant or incorrect idea /shows lack of seriousness in ideas presented/saying the same thing as in question	Complete a baseline to work out if the disease is impacting the patient over time.	Lots just say words that aren't related to the question OR repeat of the question.
1	Partly correct or incomplete /no scientifically correct terminology used/ correct without explanations	Use of relevant terms without explanation or further elaboration OR relating one idea/term correctly but having these terms explained incorrectly • Use terms that show either an understanding of symptomatic patients • No further explanation on how those terms are related	The hospital has to take a look at the big picture and work out if there is an outbreak or a disease happening.	Starting to get that there is a concept but the concept itself is wrong or it is not a developed answer.
2	A correct answer with some/no scientifically correct terminology used/ correct with explanations	Use of two or more terms that are explained correctly • Use terms that show either an understanding of symptomatic patients.	Take the necessary steps to address the population. This could include looking at beds, the number of staff allocated to certain areas, current infection control measures to ensure patient safety and effective Rx.	The idea is starting to develop however it lacks depth and is superficial in nature.

Score	Answer Type	Description	Examples	Comments
		• Scientifically correct idea of disease control and management.		
3	A correct answer with scientifically correct (Ph.D. level) terminology, elaboration, and explanations.	Reasonably complete expert explanation (which shows understanding and further relationships to relevant ideas) • Use terms that show either an understanding symptomatic patient • Scientifically correct idea of the situation	During a surge of symptomatic populations, hospitals face critical challenges in managing patient influx, ensuring adequate resources such as beds, personnel, and implementing rigorous infection control protocols are paramount. Effective triaging, optimizing treatment pathways, and maintaining robust communication channels between healthcare teams are pivotal in enhancing patient outcomes and reducing mortality rates.	They must provide a clear understanding what the spread of disease includes and provide a number of examples to demonstrate this.

Appendix L: Results of Bayesian Analysis

Table L1

Results of Research Question 1 Analysis

Bayesian Independent Samples T-Test				
	BF_{10}	error %		
Declarative Epidemics Pre Test	0.399	0.004		
Declarative Complex Pre Test	0.349	0.004		
Explanatory Complex Systems Pre Test	0.328	0.003		
Bayesian Independent Samples T-Test				
	BF_{10}	error %		
Declarative Epidemics Post Test	0.408	0.004		
Declarative Complex Post Test	0.381	0.004		
Explanatory Complex Systems Post Test	0.328	0.003		
Bayesian Paired Samples T-Test				
Measure 1		Measure 2	BF_{10}	error %
Declarative Epidemics Pre Test	-	Declarative Epidemic Post Test	56.764	0.0007
Declarative Complex Pre Test	-	Declarative Complex Post Test	8.571	0.000
Explanatory Complex Systems Pre Test	-	Explanatory Complex Systems Post Test	7.629	0.000
Bayesian Paired Samples T-Test				
Measure 1		Measure 2	BF_{10}	error %
Declarative Epidemics Pre Test	-	Declarative Epidemics Post Test	2.404	0.000
Declarative Complex Pre Test	-	Declarative Complex Post Test	14.814	0.001
Explanatory Complex Systems	-	Explanatory Complex Systems Post Test	8.868	0.000

Table L2

Results of Research Question 2 Analysis

Bayesian Independent Samples T-Test				
	BF_{10}	error %		
Near Transfer Pre Test Total	0.619	0.004		

Bayesian Paired Samples T-Test				
Measure 1		Measure 2	BF_{10}	error %
Near Transfer Pre Test Total	-	Near Transfer Post Test Total	89.856	0.0000

Bayesian Paired Samples T-Test				
Measure 1		Measure 2	BF_{10}	error %
Near Transfer Pre Test Total	-	Near Transfer Post Test Total	0.538	0.018

Bayesian Independent Samples T-Test		
	BF_{10}	error %
Near Transfer Post Test Total	0.947	0.005

Bayesian Independent Samples T-Test		
	BF_{10}	error %
Far Transfer Post	116.012	0.0000