Conceptual Framework for Designing Virtual Field Trip Games

ALSAQQAF, AMANI,ZAINALABDEEN,M

Conceptual Framework for Designing Virtual Field Trip Games

Amani Alsaqqaf

A Thesis presented for the degree of

Doctor of Philosophy

Department of Computer Sciences

University of Durham

United Kingdom

Supervisors: Frederick. Li, Ioannis Ivrissimtzis

2022

Abstract

This thesis aimed to provide designing models to explore an alternative solution for a field trip when it becomes impossible for several reasons such as the limitation of cost and time. Virtual field trip games are relatively new means to create virtual field trips in game environments through adding game aspects to learning aspects to enhance the learning experience. The simple combining of game and learning aspects will not guarantee the desired effect of virtual field trips. Theoretical and logical connections should be established to form interweave between both aspects.

This thesis proposes a designing framework by establishing three links between game design aspects and learning aspects. The three links are constructed by modelling: the experiential learning theory (ELT), the gameplay, and the game world. ELT modelling quantifies the theory into the internal economy mechanic and balances the levels of game task difficulty with the player's ability through game machinations, game modelling links the learning process to gameplay, and world modelling connects field environment to game environment. The internal economy mechanic and its components (resources, internal mechanic, feedback loop), formulating equations to define generic player's interactions and identify indicators to capture evidence of achievements via a mathematical (evaluation) model. The game modelling includes skill models to design two important high-order skills (decision-making and teamwork) and connects them to the evaluation model. The game world is modelled through defining its variables and relationships' rules to connect both environments (game and field) expanding the evaluation model. The framework is supported by essential learning theories (ELT, task-based learning, some aspects of social learning) and pedagogical aspects (assessment, feedback, field-based structure, high-order skills) and connected to the key game elements (interaction, multimodal presentation, control of choice…etc) of field-based learning along with suitable game mechanics.

The two research studies that were conducted as part of this thesis found that the designing framework is useful, usable, and provides connections between learning and game aspects and the designed VFTG based on the framework improved learning performance along with providing motivation and presence. This suggests the effectiveness of the framework.

DEDICATION

This thesis is dedicated to my parents for their care and support, to my sons Aqeel, Jaafar, and Ahmed for their patience.

Thank you all for believing in me.

Declaration

The work in this thesis is based on research carried out in the Department of Computer Sciences, University of Durham, England. No part of this thesis has been submitted elsewhere for any other degree or qualification and it is all my own work unless referenced to the contrary in the text.

The following publications were produced during the course of this thesis:

Alsaqqaf, A., & Li, F. (2019, June). Conceptual Framework for Virtual Field Trip Games. In *International Conference on Immersive Learning* (pp. 43-55). Springer, Cham.

Acknowledgements

This thesis would not have been accomplished if it was not for the support of the many people who have assisted me. I would like to acknowledge my appreciation to them all. Specifically:

My PhD supervisor; Dr Frederick Li for his direction and advice and as well as Dr Ioannis Ivrissimtzis for his assistance.

Thanks to the Saudi Arabia Cultural Bureau and Umm Al Qura University, this study would not have been feasible without their financial support.

I extend my personal thanks to:

- My friends I have met in Durham and my close friends back home for their time and encouragement.
- My parents, for their unconditional love, concern, and continued support.
- Everyone who spent their time participating in my experiments.

Finally, thanks to my sons Aqeel, Jaafar, and Ahmed without their understanding and patience this would not have been achievable.

Table of Content

List of Figures

List of Tables

List of Abbreviations

AC …………… Abstract conceptualization
AE …………… Active experiment
AES …………… Automatic essay scoring
CAT …………… Computer-adaptive testing
CBT …………… Computer-based test
CE …………… Concrete experience
D2L …………… Desire2Learn
DBS …………… Disclosure and Barring Service
DF …………… Delayed feedback
ECD …………… Evidence-centred design
EDM …………… Educational Data Mining
EF …………… Elaborated feedback
ELT …………… Experiential learning theory
FBL …………… Field-based learning
FT …………… Field trip
GBL …………… Game-based learning
ICT …………… Information and communications technology
IMI …………… Intrinsic Motivation Inventory
INV …………… Involvement
IRT …………… Item response theory
IVR …………… Immersive virtual reality
KCR …………… Knowledge of correct response
KR …………… Knowledge of result
LLTM …………… Linear logistic test model
LMS …………… Learning management system
LSI …………… Learning style inventory
NPC …………… Non-player characters
PBG …………… Problem-based gaming
PDA …………… Mobile devices
RAFT …………… Remote Access Field Trip
REAL …………… Experienced Realism

RO	....	Reflective observation
SG	....	Serious game
SP	....	Spatial Presence
SUS	....	System usability scale
TBL	....	Task-based learning
VF	....	Virtual field
VFT	....	Virtual field trip
VFTG	....	Virtual field trip game
VLE	....	Virtual learning environment
VR	....	Virtual reality
XP	....	Experience Point
XR	....	Extended reality

Chapter 1

1 Introduction

1.1 Overview

Changes in the learning process alongside constant technological developments have led to the transformation of learning and teaching into many different modes, such as e-learning, and are applied using innovative methods. Serious games (SGs) are a relatively new method for learning and training in the digital age. There are concerns about how to design SGs and to be more specific game-based learning (GBL) to ensure that they follow the required learning theories and pedagogical aspects to maximize learning outcomes. The focus of this thesis is on designing GBL to provide virtual field trip Games (VFTGs). The motivation for focusing on field trips (FTs) and fieldwork comes from the importance of field-based learning (FBL) and the continuous reduction in the number of FTs in education.

It may seem that playing games and learning are different phenomena. However, children around the world learn a lot by simply playing in the garden (learning by doing). The motivation linked to video games means that they can play for hours, and this engagement can affect academic achievement (Bawa, Watson, & Watson, 2018; Y. L. Chen & Hsu, 2020; Finn & Rock, 1997; Fredricks, Blumenfeld, & Paris, 2004). Playing video games is an optional solution to improve learners' learning (V. J. Shute, Ventura, Bauer, & Zapata-Rivera, 2009). Hence, GBL, especially those which focus on the learning and teaching of twenty-first century skills such as decision-making and problem solving, are recognized as being effective tools. However, GBL must be planned and designed based on learning theories and the application of pedagogical aspects such as assessment and feedback to ensure progress (V. J. Shute et al., 2009). The following sections will briefly discuss the most distinguished modes of learning, focusing on the essentials of GBL and FBL.

1.2 Traditional Learning

Traditional or face-to-face learning takes place in the well-known classroom setting with the learning process involving learners and teachers being physically present in the same place at the same time. There is a default assumption that traditional learning is the best mode of teaching and learning. Therefore, other modes of learning are always believed to be complementary, alternative, or less efficient. Traditional learning is a solidly established mode with advantages and limitations. The main components of the learning process can be identified as the teacher, learner, and learning content, while other aspects can be considered as tools used to support and enhance learning and teaching. The main limitation always associated with traditional learning is that it promotes passive learning where learners become listeners/receivers while the teacher is the centre of learning and the core giver of knowledge. This passive learning leads to several limitations where learners are not involved in the learning process and each learner is isolated, having to absorb the learning content by him/herself. Passive learning stimulates defining, recalling, and describing, where a given task usually has only one correct answer. Learners who have the chance to use technology can participate in the learning process by searching for knowledge and forming the learning content side by side with their teachers. The limitations of traditional learning include assessment not being adaptive in real time and that the feedback also can be delayed, which affects the final learning outcomes. In general, traditional learning is causing learners to fail to fulfil a lot of the potential opportunities offered by advanced technologies in every aspect of their lives. Technology can support learners to advance their learning to become active learners in different modes of learning such as blended learning or fully e-learning.

1.3 E-learning

E-learning is defined as any electronically supported learning and is usually linked to the instruction offered through the internet and computers. Studies have shown that the e-learning mode is at least as effective as the traditional learning mode (Faulconer, Griffith, Wood, Acharyya, & Roberts, 2018). The difference between traditional learning and e-learning is the medium via which teaching and learning are carried out. The major advantages of e-learning are flexibility and accessibility regarding time and place, which breaks down barriers enabling the learning to reach a much wider population. In addition, it enables the replay of the content of recorded lectures, notes, or even simulation scenarios such as patient cases in medical education (Dhir, Verma, Batta, & Mishra, 2017). E-learning can be classified into different types, each of which brings more benefits, such as synchronous learning, asynchronous learning, blended learning, and GBL. E-learning changes the roles of both teachers and learners; the teacher is no longer at the centre of the learning process and shifts to be a facilitator while learners become active learners. E-learning permits learners to participate in academic interactions beyond the time schedule of classes and to collaborate with learners from around the world. It brings improvements in assessment and feedback as well as collaboration and the rich availability of learning content. E-assessment involves more than electronic versions of paper tests and includes addressing learner behaviours in learning environments and providing real-time feedback. However, e-learning also has limitations such as hardware and software issues, internet connectivity, faculty support, and resistance to change and adopting a new mode of teaching. There are two main issues that restrict the potential of e-learning, which researchers should investigate in depth. The first issue is that it is easier for learners to quit the learning process; they need the motivation to continue learning. The second issue is that e-learning instructors focus more on technology and neglect learning theories, instructional design, and content development that match e-learning requirements. However, more advanced types of learning, such as gamified e-learning and GBL, would address the first issue while the second issue would be addressed by establishing frameworks for designing more effective e-learning. Finally, e-learning became a necessity during the COVID-19 pandemic and was integral to saving the lives and education of youth around the world.

1.4 Game-Based Learning

Game motivation means that players can be engaged for hours in reaching different goals, devoting their full attention to the game. Players are actively motivated to overcome their failures and to seek more tasks and challenges. The drive and elevated levels of motivation are what educators desire for their students in their learning process. Furthermore, playing games is one of the oldest and most common ways of learning. Therefore, researchers have sought to merge the content of learning with the motivation and engagement of games to provide learning environments that are more suitable for “digital natives” in order to improve learning outcomes. GBL environments allow learning to occur within a relevant context (Y. L. Chen & Hsu, 2020; Prensky, 2003). Thus, the content of learning is related to the environment in which it can be applied and practised (Fu, Lin, Hwang, & Zhang, 2019; Van Eck, 2006). Regarding the application of game playing in education, a number of questions can be asked: Is GBL really engaging and effective? Why, when, and under what circumstances should it be implemented?

Generally, games provide physical or mental competition with the purpose of entertaining players, who engage in various challenges on different levels and according to certain rules. Video games are played on a computer or any electronic device (e.g. mobile phone and tablet) (Zyda, 2005). Games created to serve a specific purpose (learning, training, or changing behaviour) in any field such as health, education, government, or social connection along with fun and entertainment are called SGs. GBL is the method and practice of learning through the use of games. GBL can utilise digital or non-digital games as well as SGs or entertainment games. GBL configured in SG aims to enhance the purpose associated with the SGs. Regarding digital games, GBL can be achieved in two ways: educators and game developers either collaborate to build games or use commercial off-the-shelf games; each way has its advantages and disadvantages. It is important when designing and implementing games for learning to not take the fun out of it. A learning process that uses games as a tool of education, training, or even assessment benefits from the fun and motivation effects. Therefore, “academizing” (Van Eck, 2006) of educational games should be avoided. Also, reducing the game design to one or two game elements/mechanics would limit the learning effect of GBL (Liao, Chen, & Shih, 2019); there is a need to create a balance between pedagogical and engagement components.

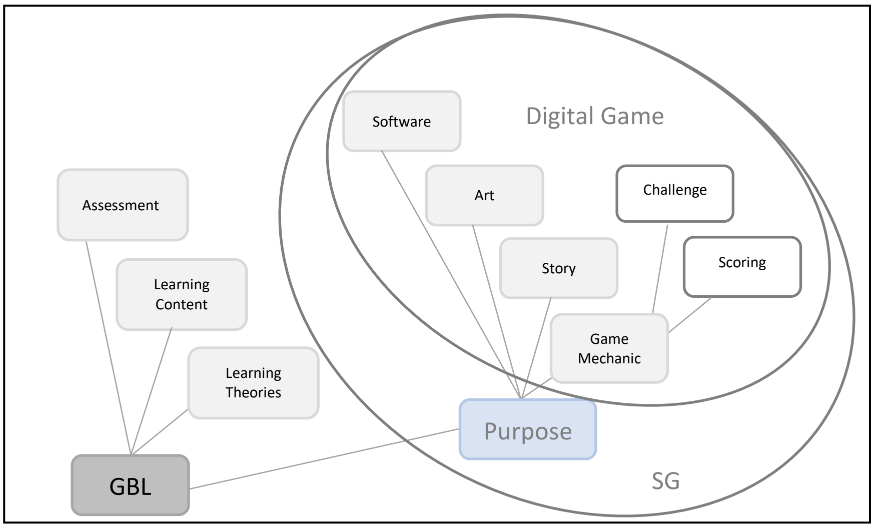


Figure 1.1: GBL Components.

Game playing has different taxonomies: there is a very basic classification of digital games based on one criterion such as domain, genre, market or learning outcome; or the classification can include complex games containing more dimensions, namely, domain and purpose (Djaouti, Alvarez, & Jessel, 2011; Djaouti, Alvarez, Jessel, Methel, & Molinier, 2008). An effective SG consists of a combination of video game components and a specific purpose else fun. There are four main elements involved in building a video game, as displayed in Figure 1.1: a story, which introduces fun and engagement; art to create graphics and interface characters; game mechanics to create challenges and link them to the story and art; and lastly software to facilitate programmatic and technical aspects such as coding the game mechanics, story, and interface graphics. Adding a specific purpose to the entertainment video game would produce a SG. However, when all this is governed by pedagogies/learning theories to enhancing the purpose with rich learning content would improve the SG so that it provides GBL.

Figure 1.1 shows the relation between the components of digital games, SGs, and GBL. Educators have to work with game designers to specify learning objectives, certain skills, and learning theories to ensure the effectiveness of GBL. Moreover, analysis and reporting of player progress are mentioned in the literature. This mechanism could be applied to provide assessment and feedback. Scoring and feedback is an essential part of games, keeping players motivated and guiding them to their next goal.

In summary, a number of important points should be highlighted:

- In designing GBL, the story component provides the fun and the motivation to enhance the learning process. This means that the fun components come first and then the pedagogy, in order to avoid repeating the failure of edutainment games, which focus on learning theories and practice with an absence of the fun component. These games were not engaging and did not achieve their purpose of motivating learners to learn while playing.
- Building GBL requires two teams of experts working together. The first team consists of educators, instructional designers, and subject teachers, and the second team includes game designers and developers addressing all tasks related to graphics, code implementation, networking, and story. Each team builds part of the GBL and ensures that the storyline and pedagogies fit together to provide the required learning environment. Conceptual frameworks can help game designers and developers in absence of educators to bridge the link between learning aspects and game aspects.
- Researchers study games from two angles: they need to know what games provide and how games support learning. Other researchers are interested in developing the instructional design of the learning process in games such as the game-based feedback mechanism.

1.5 Field-Based Learning

FBL is learning through first-hand experience, outside the restrictions of classroom walls (Lonergan & Andresen, 1988), and a field trip represents a unit or lesson of the curriculum. Fieldwork is the set of activities carried out by learners in the field to learn by doing, while the field is where learning takes place outside of the traditional classroom. Fieldwork experience is essential in different subjects such as geology, ecology, zoology, science, and civil engineering. Ryle (1949) distinguished two types of knowing: knowing “that” and knowing “how”. Knowing “that” means acquiring knowledge about something, while knowing “how” to put knowing into action. This is a major difference between fieldwork and other learning activities, where learning knowledge and theories are linked to practice in the field (Fedesco, Cavin, & Henares, 2020).

FBL differs from traditional classroom learning in a number of points: the place, methodology of learning, and assessment. The place could be anywhere outside the classroom, starting from the school yard, museum, park, or going to another country/continent. The methodology of learning in FBL requires first-hand experience learning, and a sense of reality, while the fieldwork environment forms a learning resource. Assessment of learning outcomes that results from fieldwork always raised questions of how, what, and when to evaluate. In addition to the validity and reliability of the assessment method, the most well-known assessment examples are writing in notebooks or performing a presentation. In recent times there has been a reduction in the number of FTs and fieldwork across all learning levels, not only in the UK but around the world. There are numerous reasons for this decline. For example, FTs and fieldwork may have become less common due to time and financial limitations, an increased number of learners, and safety issues. In some cases, FTs are limited to groups of learners who have achieved high grades, so going out into the field functions as a reward.

1.6 Motivation

This section will explain why the research into designing VFTGs based on a conceptual framework is worth studying. Advances in technologies have remarkably improved the accessibility of technological devices. In conjunction with emerging gaming concepts, digital games have become a popular method for developing applications. Learning in particular is considered as an application of gaming with strong potential, and comprehensive learning games can be easily created by utilising available game engines and platforms.

Digital games could provide an environment to teach and train learners through different methods of teaching. Gaming is adapted to create promising learning environments to attract learners, especially digital-age learners. The characteristics of games can enrich the learning process and improve learning outcomes. GBL is the mode of learning that utilises games to serve learning purposes and nowadays the focus is on digital games. Researchers have studied the benefits of GBL and developed ways to guarantee its effectiveness. GBL is considered as a type of e-learning which has become a part of learning in learning institutions around the world and a necessity due to the global COVID-19 pandemic.

Technologies enhance learning and teaching in general and transfer specific learning modes to advanced dimensions such as FBL. FBL provides learners with hands-on experience to link theories and factual knowledge with practice in real-world situations. E-learning such as web-based multimodal FTs is applied as an alternative to FTs, which are in decline. Some of the reasons for this reduction are the limitations of time and costs, the continuous increase in learner numbers, learner safety, especially disabled learners, and recently the global pandemic. web-based multimodal FTs provide an alternative to physical FTs but yet remain limited although advanced technologies are capable of delivering more efficient options.

Digital game environments can be manipulated to create FT environments and game characteristics such as role playing, storytelling and more can be utilised to improve FBL through GBL. VFTs in game environments create a virtual field (VF) to provide hands-on experience and at the same time give learners a feeling of presence, a “sense of being there”, and many other aspects such as motivation and engagement. However, combining GBL and FBL requires a theoretical and conceptual framework to guide the design and implementation to ensure the effectiveness of VFTG. Educators/ game designers can benefit from a conceptual framework that ensures a balance between fun and learning by linking learning theories to game elements and mechanics.

1.7 Problem Statement and Aim

It is necessary to explore the design of GBL, particularly due to the rapid involvement of technology in education. The design of GBL must be reconsidered specifically in relation to VFTGs. Unlike prior research studies which mainly focus on enabling the design of general GBL, this research study addresses the need for a better process to design FBL as GBL. Field trips and fieldwork form an essential method of teaching and learning although they have been reduced around the world for different reasons, which raises an urgent need for an alternative solution. As such, this thesis investigates the possibility of enhancing the design of VFTGs as a promising alternative with a more comprehensive conceptual and theoretical framework.

This study has been conducted in order to address the following questions:

Q1) What is the current intersection between GBL and FBL?

Answered in Chapter 2.

Q2) What are the possible strategies to connect FBL aspects to game design aspects?

Answered in Chapters 2, 3, and 4.

Q3) To what extent can the proposed conceptual framework facilitate and build a linkage between FBL aspects and game design aspects?

Answered in Chapter 5.

Q4) To what extent can the proposed conceptual framework improve the learning process of virtual field trip games?

Answered in Chapter 7.

The research aims to improve the design process of VFTGs, in order to facilitate teaching and learning in VFTGs. In this study, a conceptual framework will be developed and evaluated. The proposed conceptual framework contributes to building up a linkage between theories and principles of FBL and game design techniques.

1.8 Contributions

The contributions and originality of this research lie in the following points:

- Three links between game design aspects and learning aspects of FBL are developed to facilitate the design of VFTGs for game designers/educators. The links are built as models to quantify learning aspects based on the game mechanic (internal economy), machination framework (balancing levels of difficulty with levels of ability), and identifying performance indicators:

  - ELT modelling (mapping the experiential learning theory (ELT) into the internal economy mechanic).
  - Game modelling links the learning process to gameplay through (matching scheme, task modelling, evaluation model, and skill modelling).
  - World modelling (connecting the field environment to the game environment).
- Supporting the design of in-game assessment through a mathematical model of evaluation that defines indicators of players' interactions. The expected series of interactions are modelled to capture and track the players' performance evolution with a generalisation and standardisation of assessment based on the defined indicators. The same model is extended to cover the game environment (world modelling) in relation to in-game assessment.

1.9 Thesis Outline

The remainder of this thesis will be structured as follows:

Chapter Two (Literature Review): investigates and reviews several major issues related to GBL and FBL, specifically: computer-supported field trips, designing VFTs as GBL, a conceptual framework for designing GBL and FBL, and analysing different methods of evaluation and feedback in learning with a focus on assessment in educational games. Also, it includes a background review of educational games about geography and specifically volcanoes.

Chapter Three (Research Methodology): presents the methodology that has been followed to perform this research and briefly explains the design of the two studies (preliminary and main) that have been conducted. In addition, the chapter outlines the research elements such as data collection instruments and statistical analysis tests.

Chapter Four (The Conceptual Framework): introduces a new framework built on learning theories and pedagogical concepts to provide connections to game design with the aim of facilitating the design of VFTGs.

Chapter Five (Preliminary Study): explains the first research study in detail by presenting the study questions and the steps followed to design and conduct the study. Also, the chapter presents the statistical analysis tests, results, and findings along with the actions taken based on the findings.

Chapter Six (Prototyping): describes how the prototype (Island of Volcanoes) was designed based on the proposed framework and implemented. The chapter includes a discussion of the learning aspects, the establishing of the three links, and the implementation process.

Chapter Seven (Main Study): discusses the study design utilised and the research instruments used along with some difficulties. Research questions are translated into hypotheses and the relationships between dependent and independent variables are determined. In addition, data analysis is explored, starting with the validity and reliability of the study instruments. Also, the chapter presents the quantitative methods applied to analyse the collected data, together with the results. Finally, the chapter presents the findings and reflects on the proposed framework.

Chapter Eight (Discussion and Conclusion): presents a discussion, general conclusions, limitations of the research, and directions for future research.

Chapter 2

2 Literature Review

2.1 Introduction

This chapter focuses on reviewing the literature, examining different concepts that will help in forming a general understanding of some areas and analysing other areas in depth which will help to answer the first question of this research that was identified in chapter 1. The topics included in this review are as follows: the relationship between GBL and FBL, technology-associated field trips, frameworks for designing SGs/GBL, e-assessment tools, pedagogical aspects that concern feedback, and geographical video games. Selected papers from these areas will be analysed and explored to identify challenges and limitations. For the purpose of this research, a search of the literature was performed in the following databases: Web of Science, IEEE Xplore, ACM library, Taylor & Francis, Google Scholar, JSTOR, and QTI. Keywords were specified and the operators "AND" and "OR" were utilised to delimit the search criteria. The following criteria were used: (game-based learning OR serious game), (assessment OR online assessment OR e-assessment), (automatic feedback OR feedback AND (game-based learning OR online learning)), (computer-aided OR technology supported OR remote access AND (field trip OR virtual fieldwork)), (game design framework OR educational game framework).

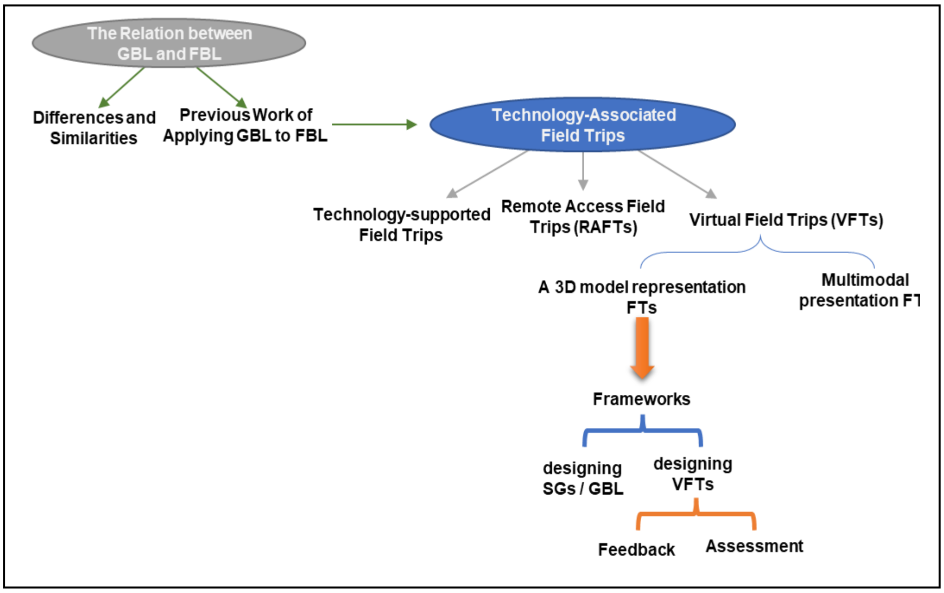


Figure 2.1: Literature review roadmap.

The selected papers were deemed significant if they met the following criteria:

- The papers discussed in depth one of the topics mentioned above.
- The papers were published between 1999 and early 2022.

This chapter is organized into sections as displayed in Fig. 2.1, each of which covers one topic: the relationship between GBL and FBL, technology- associated field trips, frameworks for designing SG/GBL, e-assessment tools, feedback, and geographical video games.

2.2 The Relation between GBL and FBL

Virtual field trips (VFTs) in the game environment specifically form a promising area of research. Therefore, the starting point is to analyse both GBL and FBL in order to understand the relation between them, if any. The similarities and differences regarding the structures and learning theories applied to both modes of learning (GBL and FBL) are explained in the following section. Then, the findings from previous studies on combining them together to enhance learning are presented.

2.2.1 Differences and Similarities

The current research involves combining these two learning mods in order to enhance learning. Therefore, in this section the similarities and differences between GBL and FBL will be outlined. These are two different modes of learning but share the need for three components concluded from reading and analysing the literature: a) characteristics of learning environment settings, b) pedagogical aspects, and c) activities.

Despite the similarity in basic components between the two methods of learning (GBL and FBL), the essential difference is in their purpose and goals, which distinguish the components of each learning method. a) The characteristics of learning environment settings are important to FBL because the learning environment represents a fundamental source of learning content. Learners have to observe and analyse the field (forest, museum, theatre, or zoo), which is the learning environment. In GBL, learning content is embedded in a game environment and the main purpose of this environment is to provide motivation and engagement. Environment settings are used in GBL to engage and immerse learners in embedded learning materials, while FBL considers the learning environment as learning content in itself.

Regarding designing a digital GBL, it has to be modelled carefully to provide more than motivation by applying learning theories, which is the second shared component. b) The pedagogical aspects are essential in both GBL and FBL. Field trips (FTs) and fieldwork are well-known methods of learning, and a huge number of studies in the literature concentrate on studying their effectiveness and analysing their learning outcomes (DeWitt & Storksdieck, 2008). In contrast, learning theories for FBL are limited and hard to find in the literature. The most known and used theory in FBL is Kolb’s Experiential Learning Theory (Behrendt & Franklin, 2014; Bonello, 2001; Dummer, Cook, Parker, Barrett, & Hull, 2008; Dunphy & Spellman, 2009; Healey & Jenkins, 2000; Scott et al., 2012). In addition, reflection (Ladyshewsky & Gardner, 2008; McGuinness & Simm, 2005; Scott et al., 2012) is an important part of experiential learning theory, and deep learning is an expected result from reflection (Hill & Woodland, 2002). Furthermore, collaborative learning (Bartholomai & Fitzgerald, 2007; Cohn, Dooley, & Simmons, 2002; Jung, Sainsbury, Grum, Wilkins, & Tryssenaar, 2002; Mason, 1998), problem-solving (Bradbeer, 1996; Perkins, Evans, Gavin, Johns, & Moore, 2001), and critical thinking (Greene, Kisida, & Bowen, 2014; Gutwill & Allen, 2012) is applied to FBL.

There is a huge mass of literature on GBL that references applying pedagogical aspects and theories. Li and Tsai (2013) reviewed GBL from 2000 to 2011 and reported four dimensions of theoretical foundation: theory, model, approach, and principle. The scope of the current research cannot cover much of the detail about the pedagogical aspects applied to GBL but will suffice by examining learning theories that have been explicitly referenced in studies on GBL. Flow theory is utilised in GBL explicitly and implicitly for enhancing engagement, motivation, and the potential to enable effective learning (Barzilai & Blau, 2014; Brom et al., 2014; Chang, Wu, Weng, & Sung, 2012; Hamari et al., 2016; Kiili, 2005; Liu, Agrawal, Sarkar, & Chen, 2009; Procci, Singer, Levy, & Bowers, 2012). Other learning theories and aspects used in GBL are: constructivist learning theory (Pivec, Dziabenko, & Schinnerl, 2004), cognitive load theory (Annetta, Minogue, Holmes, & Cheng, 2009; Spires, Rowe, Mott, & Lester, 2011; Wrzesien & Raya, 2010), activity theory (Barab, Thomas, Dodge, Carteaux, & Tuzun, 2005; Spires et al., 2011), and Kolb’s Experiential Learning Theory (Kiili, 2005; Wrzesien & Raya, 2010). GBL also uses problem-solving (L. Miller, Moreno, Willcockson, Smith, & Mayes, 2006; Pivec et al., 2004; Spires et al., 2011) and collaborative learning (Pivec et al., 2004; Sung & Hwang, 2013).

FBL and GBL require learning theories and pedagogical aspects that ensure effective learning and enhanced outcomes. Experiential Learning Theory (ELT) is a well-established theory utilised in both FBL and GBL, as well as collaborative learning. FBL has a specific structure: pre-, during, and post-field trip (Behrendt & Franklin, 2014; Bitgood, 1989; A. M. Kamarainen et al., 2013; Kent, Gilbertson, & Hunt, 1997). Studies show that this structure contributes to creating an effective fieldwork experience and learning outcomes (Davidson, Passmore, & Anderson, 2010; DeWitt & Storksdieck, 2008). This structure is not applied to GBL unless it is GBL for FTs, or a mobile game used to enhance a physical FT (Chih Hung Chen, Liu, & Hwang, 2016; Sommerauer & Müller, 2014).

The last component is c) learning activity/task. FBL activities are manipulated by: time and location (Kent et al., 1997; Lonergan & Andresen, 1988), and participation and autonomy (Kent et al., 1997). Some could argue that time is a variable of any learning task, which is definitely the case, but combining time with location is unique to FBL. In traditional learning, the location is always the classroom: it is a constant variable. From the distance-learning perspective, the location can be anywhere but this variability (being at home, public library, or even a park) would not and should not add value to the learning task or outcome. On the other hand, fieldwork activities are built upon different locations, which are assumed to provide the necessary learning content and context. The importance of location in FBL activities leads to the concept of authenticity. The FBL activities also differ in degree of observation or participation and dependency on teacher guiding or self-guiding.

The nature of FTs and experiential learning provides an authentic learning environment (Behrendt & Franklin, 2014). Nevertheless, FTs should be supported with authentic tasks in order to ensure completely authentic learning. On the other hand, authenticity in GBL concentrates on the setting or presentation (Arnab et al., 2015; Yusoff, Crowder, Gilbert, & Wills, 2009). In GBL, authenticity is used to increase motivation and immersion (Papanastasiou, Drigas, & Skianis, 2017; Westera, Nadolski, Hummel, & Wopereis, 2008), which can be achieved with authentic content such as in *Pac-Man*. The authentic presentation should be used as it is needed to support the content in order to achieve the learning/training goals as in *America's Army: Rise of the Soldier*. Authentic presentation is essential for VFTs because it provides a VF to learners that should replace the physical field with some level of fidelity. I would argue that another type of authenticity should be considered in designing GBL/FBL, which is authentic behaviours (Zielke et al., 2009).

2.2.2 Previous Work of Applying GBL to FBL

GBL contributes to FBL in different ways. First, some mobile learning games are designed to be used during physical FTs to help learners by guiding them (Chih Hung Chen et al., 2016; Sánchez & Olivares, 2011), encouraging exploring and engagement (Atwood-Blaine & Huffman, 2017), or collecting data (Kingston et al., 2012). The main purpose of these in-field mobile games is to support the learning process in physical FTs and they do not provide the complete FT experience by themselves. Second, there are a number of digital VFTs implemented in game engines such as Unity 3D (Bourke & Green, 2016) and Unreal (Harrington, 2009) to achieve the desired authenticity of field presentation. For example, Harrington (2009) provides a data-based visualization model of objects and landmarks of a field (Trillium Trail in the U.S.) and the user can navigate through the virtual world and find different types of information (text, photographs) about 3D model objects. However, it is just a high-fidelity virtual presentation of a physical location without any game elements or learning activities. Third, few VFTs have been designed and implemented to provide GBL (Argles, Minocha, & Burden, 2015; Bursztyn, Walker, Shelton, & Pederson, 2017; Georgiadi et al., 2016). In fact, these virtual field trip games (VFTGs) are more like gamified VFTs than SG or GBL. The implementations pick a handful of game elements to apply and miss out the gameplay. For example, Bursztyn et al. (2017) used a learning theory and framework (Kiili, 2005) to link game elements to learning; however, there is no gameplay, narrative, or interaction, and a lot of pop-up multiple-choice questions which interrupt the gameplay. Georgiadi et al. (2016) implemented a VFT educational game for children and used a framework (Hunicke, LeBlanc, & Zubek, 2004) for designing video games without any consideration of learning elements such as learning theories, and the authenticity of presentation is also ignored in my opinion. Argles et al. (2015) created a good VFTG called *Virtual Skiddaw* to support distance learning. Again, the VFTG does not provide gameplay or learning theories, and the interaction is reduced to clicking to show information, mostly text. These examples of VFTG show the need for a framework to support this type of learning experience to link FBL elements to game elements. Some of these missing links are the structure of FTs (pre-, during, post-) that should be interweaved into the game design; high-order skills development which have been almost completely ignored in the previous examples of VFTG; and game elements (gameplay, narrative, interaction) that should be applied to support learning.

2.3 Technology-Associated Field Trips

A review was conducted of more than the last 10 years (2008–2022) of research. Some of the papers in the search results were excluded from the review due to them adopting a misleading meaning of fieldwork. Fieldwork can be defined as a number of activities carried out by learners where they are involved in first-hand experience; the field is where learning takes place outside the traditional classroom, and a field trip represents a unit or lesson of the curriculum. The literature shows that technology is associated with FBL in three ways, which can be categorized as follows: technology-supported field trips, remote access field trips, and virtual field trips. The literature was analysed based on three criteria: type of technology used in the study, learning theories, if any, applied, and framework followed to create the FT experience.

Technology-supported Field Trips use smartphones, iPads, and augmented reality in physical FTs to enhance the learning experience (Dyson, Lawrence, Litchfield, & Zmijewska, 2008; Hsu & Chen, 2010; A. M. Kamarainen et al., 2013; Kayalar & Balcısoy, 2008; Kingston et al., 2012; Shinneman, Loeffler, & Myrbo, 2020; Welsh & France, 2012; Welsh et al., 2015). The main goal is to support the learning process in FTs by using technological devices to collect data such as photographs, videos (Efstathiou, Kyza, & Georgiou, 2018), and GPS coordinates (Dyson et al., 2008). Technology is also used to guide learners in their FTs (Chih Hung Chen et al., 2016; Hsu & Chen, 2010; A. Kamarainen, Reilly, Metcalf, Grotzer, & Dede, 2018; Shinneman et al., 2020), promote interaction in specific physical locations (Chin & Wang, 2021; Efstathiou et al., 2018; A. M. Kamarainen et al., 2013), and provide more information (text, audio, video, 3D models) (Kayalar & Balcısoy, 2008; Verdes, Navarro, & Álvarez-Campos, 2021). Mobile technologies can be used for a range of purposes, extending from facilitating a single exercise in a real FT to helping develop transferrable skills such as the collection of spatial data (Kingston et al., 2012). Technology-supported field trips seem to focus on using learning theories to create opportunities for collaboration between peers in the field sites, and to encourage interaction with elements in the field (Dyson et al., 2008; Kingston et al., 2012). Mobile apps can be used on mobile devices to facilitate learning activities in FTs. For example, Twitter can be used as a tool of discussion, the camera app helps to capture techniques and to keep a record of observation, while Geospike can be used to create maps (Welsh & France, 2012; Welsh et al., 2015). However, the use of mobile devices in real FTs has a main drawback: the possibility of poor internet connectivity in some locations (Welsh et al., 2015). A final point to consider when using mobile devices in FTs is that the aim should be to transform knowledge and learning rather than just substitute traditional methods of learning with new sophisticated technologies (Thomas & Munge, 2017). The literature shows that using mobile devices in FTs focuses on a few learning aspects, such as encouraging interaction and collaboration or supporting self-guided learning. There is no evidence in the literature that frameworks are used to support the development or use of technology in FTs.

Remote Access Field Trips (RAFTs) aim to create opportunities for studying in rural field sites, reducing costs and, most importantly, supporting disabled learners to have an equal education (Collins, Gaved, & Lea, 2010; Palaigeorgiou, Malandrakis, & Tsolopani, 2017; Stephens, Pallant, & McIntyre, 2016; Stokes et al., 2012). RAFTs require the teacher and maybe a group of learners to actually be present at the field site while the other learners follow the activity by watching a live podcast. Further information and tools are provided to learners in remote sites, such as magnified images, maps, campus information, and communication tools (text chat, telephone call) (Collins et al., 2010; Stephens et al., 2016). Communication tools encourage inquiry learning and collaboration between the remote learners and the teacher and physically present learners. The major difficulty of this type of FT arises from technical issues that may interrupt the learning process, such as the loss of the internet connection. Also, technical support is required during the whole session of the remote field trip to assist learners with technology problems. However, regardless of the support and technologies used to create fieldwork experience, learners have reported that they could not feel a sense of “being there” (Stephens et al., 2016). The nature of RAFTs limits learning outcomes by focusing on a few objectives at a time, such as observation, encouraging peer interaction or inquiry learning, and will not be very effective in developing learning skills (Palaigeorgiou et al., 2017). RAFTs do not require a design framework, in my opinion because they are supposed to follow the teacher in the field and be designed by the teacher. In fact, live and recorded broadcasts are used as VFTs and made available to learners via the internet (Australian Government Office for Learning and Teaching; Discovery Education; Seifan, Dada, & Berenjian, 2020). Seifan et al. (2020) presented an example of VFT recorded video for construction sites as a part of a highway and transportation engineering course. There is no mention of any pedagogical aspects or learning activities that are applied to implement the VFT as it is just a recording version of a FT. However, students found this method of FT is less important than actual FT and lectures. Also, they found it less motivation to select careers. The COVID-19 pandemic encouraged universities to create innovative methods of teaching as Aleman, Duball, Schwyter, and Vaughan

(2021) present a recorded RAFT that is enhanced with a take-home field kit. The RAFT is part of a soil science course, and the kit includes soil samples from different field sites among other tools. The students have to watch the recorded RAFT and then analyse the soil samples and answer some questions on a sheet. The students appreciated the field kit and found it helpful. However, the students reported the lack of communication with peers and the instructor. There was not enough explanation of the process of designing the RAFT or the learning activities else stating the learning objectives. There is not much information in the literature about the methodologies or learning aspects of RAFTs. However, learning tasks are usually provided in separate files to complete, such as observation worksheets or reflection questions.

Virtual Field Trips (VFTs) are an alternative option to a real FT used to support experiential learning in many subjects. VFTs are conducted virtually where the presentation of the field is brought to students via different technologies such as web or gaming platforms. VFTs have two different representations in the literature: 1) multimodal presentation and 2) 3D model representation.

<u>Multimodal presentation FT</u> is considered virtual as it provides multimodal information via the internet (Cheng & Tsai, 2019; Green & Bojar, 2010; Lenkeit Meezan & Cuffey, 2012; Paleontological Research Institution; Stott, Nuttall, & McCloskey, 2009; University of Southampton). This kind of VFTs usually supports experiential learning in general without applying specific learning theories. In addition, designing VFTs follows a specific structure or template, which is populated with images, maps, and information. VFTs create the fieldwork experience by using technological elements such as Google Maps (map, terrain), photographs (panoramic, airborne, satellite) of the field site, video clips, and Google Earth in addition to verbal information (descriptions, academic papers, exercises) displayed in a web page. For example, Lenkeit Meezan and Cuffey (2012) produce VFTs for geoscience classes that can be accessed via computers or smartphones. Each VFT has a specific structure: overview, which is a summary of the real FT; observations (photographs and maps); process, which consists of descriptions and discussions; application as an exercise; and further study materials, which are all presented in a webpage. Sight is the only sense used to simulate the fieldwork experience, while background sound (birds, wind, and water) can be added from the real field sites. Problem solving and an inquiry approach are supported by a worksheet of questions to guide learners in their observation and analysis. However, assessment of learning outcomes has shown that learners successfully completed knowledge-based tasks and struggled with higher-order skills

such as analysis. Learners' feedback shows interest in photos that show actual learners doing activities at the real field sites (Lenkeit Meezan & Cuffey, 2012). VFTs can be presented to students via web platforms such as Google Expeditions (Cheng & Tsai, 2019), or ThingLink (Kenna & Potter, 2018), Google Earth (D. D. Gregory, Tomes, Panasiuk, & Andersen, 2022). Cheng and Tsai (2019) create VFT and focus on two learning concepts (observation and cook's tour field trip) but without any guidance on how to achieve these aspects while Kenna and Potter (2018) talk about planning the VFT based on ELT, however it was limited. D. D. Gregory et al. (2022) present two different types of VFT: multimodal VFT and 3D model of a field site. The paper did not provide any explanation of pedagogical or designing aspects else pointing to the learning objectives and the data that used to create the high-level presentation. Also, all the tasks are explained via YouTube videos and performed also outside the VFT environment.

It can be concluded that there is a need for engagement and immersion within VFTs; learners want to interact in the same way as physically present learners do in FTs. This type of VFT provides a good alternative to a real FT with a number of limitations that can be overcome by today's technology. The literature shows that usually no frameworks are used to design and develop this type of VFT; instead, a simple webpage structure is used (Lenkeit Meezan & Cuffey, 2012), the teacher guiding (Cheng & Tsai, 2019), or a PowerPoint or Prezi template (Paleontological Research Institution). The strength of this type of VFT comes from the supporting materials used, such as exercises and reading materials, which can enhance the learning experience.

<u>A 3D model representation</u> of space with a different level of authenticity and complexity to a real-world field, lab, or museum is a type of VFT (Argles et al., 2015; Ashfield, Jarvis, & Kaduk, 2010; Burden et al., 2017; Calvert & Abadia, 2020; Getchell, Miller, Nicoll, Sweetman, & Allison, 2010; Mathews, Andrews, & Luck, 2012). VFTs should provide learners with more than a sophisticated presentation, such as activities that mimic traditional fieldwork tasks (exploring, collecting data, and analysing). This type of VFT should aim to create the whole FT experience in the virtual world by integrating elements such as pedagogical aspects, learning content, and exercises, along with engagement, immersion, and motivation, delivered to teachers and their learners. However, the literature is full of examples where the design of VFTs lacks input from learning theories. For instance, Sedimentary Rocks: The Salona Formations (Zhao et al., 2020) creates 360° images in Unity 3D and during the learning experience, learners are asked to observe and take notes and encouraged to click icons to access

supplementary information. However, the assessment is done during the experience but separated from the VFT. This work has some potential but lacks basic elements (learning theories and design concepts).

EQUILIBRIUM game (Neuwirth, 2020) is an example of educational game designed to provide a VFT to a fictitious island that teaches students about systems thinking and modelling through the human-environmental system. The VFTG is designed as a point-and-click field trip where the player plays a role of a bird to explore the island and find out the natural system that led to the disappearance of humans. When the player clicks on some items, information will be displayed, or an action will be triggered. There was no mention of utilising a model or framework to guide the process of designing or even a discussion of any learning theories. However, two points were discussed in the study (prior knowledge of player/prior activities and rewording system) that showed the need for a framework to enhance the designing of VFTGs and overcome these limitations. The proposed framework of this thesis guides to utilise the designing of VFTG's activities based on the prior knowledge of players, a mechanism to prepare the player for the VFTG in a pre- phase as a part of the second link (see Chapter 4) of the framework, and also an explanation of the most essential game elements such as rewording system and how to employ them to enhance the learning experience.

(Needle, Crider, Mooc, & Akers, 2021) present two VFTs in a videogame environment with virtual tools such as a compass, ruler, and jetpack. The designing of both VFTs follows a framework for designing VF experiences (Atchison, Burmeister, Egger, Ryker, & Tikoff, 2020). Unity 3D is used for the implementation, however none of these VFTs utilised any game mechanics or game elements.

The number of attempts to provide VFTs in game environment are increased in respond to the global pandemic and its effect on education. However, most of these attempts do not utilise the game mechanics/elements to support the FBL (Harrington, Bledsoe, Jones, Miller, & Pring, 2021), employ one or two elements such as story and minimum interaction (clicking) (Chan, Chan, & Fong, 2020; Neuwirth, 2020), or collaborative discussion with peers (D. D. Gregory et al., 2022). Focusing on high-level of fidelity/presentation and factual knowledge could be the reason for ignoring the value of employing game mechanics/elements to design VFTs even though game engines are used for the implementation. The designers and educators seem to try replicating the presentation of actual field sites and forgetting the unlimited possibilities of learning activities that could be designed via game mechanics/elements.

There is a real urgency for innovation in designing and developing VFTGs for a number of reasons, including reduction of time and costs, increased number of learners, and ensuring learner safety. When designing and developing VFTs, teachers should be able to create a unique virtual experience that is more suitable for their learners' needs (Bellotti, Berta, De Gloria, & Primavera, 2010; Burden et al., 2017). From the literature, it is evident that VFTs should involve more than using the most advanced technology. They are a way of adapting technology to enhance learning. Therefore, learning theories and pedagogical aspects need to be linked and embedded into the used technology.

2.4 Frameworks for Designing GBL/VFTGs

This part of the literature review covers papers from 2005–2022 focusing on the analysis of frameworks for designing VFTs or GBL/SG. Some papers were excluded as the presented frameworks considering the design of virtual reality learning environments that do not provide experiential learning or VFT experiences (Hajirasouli, Banihashemi, Kumarasuriyar, Talebi, & Tabadkani, 2021; Marougkas, Troussas, Krouska, & Sgouropoulou, 2021; Scurati, Bertoni, Graziosi, & Ferrise, 2021). This section is summarized in Table 2.1 and concluded with recommendations.

Norman (1993) mentioned that essential elements of the learning environment can, generally speaking, be found in games. Playing and learning seem like two different things – one provides fun and the other requires studying – but at the same time people learn during daily life while playing in the backyard with friends. However, fun is not the magic ingredient for successful learning and teaching. Different game elements can lead to different learning experiences and support different learning outcomes.

The following frameworks support designing SGs and GBL. Some of these frameworks focus on specific elements of game design and build the framework around them, such as activities (Chew, 2017), narrative (Z. H. Chen, Chen, & Dai, 2018), or generating scenarios for training (Luo, Yin, Cai, Zhong, & Lees, 2017). Alexiou and Schippers (2018) discuss a conceptual framework that mostly supports engagement and motivation and analyses the learning process and other game elements around them. However, there is no clear guide to help game designers/educators to design GBL. Also, some frameworks support specific subject areas such as environmental learning (Fjællingsdal & Klöckner, 2017), health purposes (Ushaw, Davison, Eyre, & Morgan, 2015; Ushaw, Eyre, & Morgan, 2017), or a particular genre of digital game (massively multiplayer online role-playing games) (Eseryel, Guo, & Law, 2012).

Some frameworks focus on issues related to GBL design, such as designing assessment (Abdellatif, McCollum, & McMullan, 2018; Graafland et al., 2014; Mitgutsch & Alvarado, 2012), difficulty adjustment of tasks (Shena, Sitohang, & Rukmono, 2019), and reducing the complexity of SG design Westera et al. (2008).

More comprehensive and general-purpose frameworks of GBL design are considered in some detail. De Freitas and Jarvis (2006) produced a framework with four dimensions: context, learner, representation, and pedagogy. The framework includes essential dimensions but with a very general discussion and explanation to facilitate designing educational games. For example, the pedagogy dimension lacks important elements of the learning process such as feedback and assessment. Also, the context dimension discusses the context of playing the game in class or in any location but does not cite the learning context within the game world and its relation with, and effects on, the learning process. Yusoff et al. (2009) present a richer conceptual framework that covers more game attributes and learning elements. However, the connections between the framework's components are foggy and unclear. It resembles a sequential method of input-process-output, which seems to be the opposite of the iterative process of game design.

The experiential gaming model (Kiili, 2005) can be considered as a designing model of VFTG because it is mainly based on ELT, which is frequently applied to facilitate FBL. The model aims to integrate the learning theory within game design to design GBL. Flow theory and its main characteristics are emphasized to enhance the positive experience of players, which means that they are more engaged with learning. The model applies a specific learning theory, ELT, and focuses on the importance of clear goals, immediate feedback, and challenges based on players' skills. Two main elements are not included in this model: social interaction and assessment. Focusing on one learning theory could be considered a disadvantage which limits game designers/educators to build the game based only on this theory. In addition, the model emphasizes a few game elements (story, game balance, appearance) and their relation to education. Kiili (2007) created another model problem-based gaming (PBG) (Kiili, 2007) – that can be considered as an addition to the experiential gaming model. The main addition is the use of problem-based learning (PBL) in designing games along with collaboration, authenticity, and learning by doing. In general, the model concentrates on describing the learning process in gaming and ignores the integration of pedagogical theories into game elements, especially engagement and motivation. This model cannot stand alone to design GBL but could be part of a richer framework used to support PBL; the model also does not discuss issues related to individual/multiplayer learning in PBG.

The Game Object Model II (GOM II) (Amory, 2007) is a developed version of the Game Object Model that uses the OOP system paradigm as a metaphor to help in understanding complex concepts of game design. Social interactions and learning tasks represent the major developments in this version. The model does not focus on one learning theory and instead applies different pedagogical theories and relates them to game elements. However, the model does not emphasize the value of assessment and feedback, with this being mentioned only once without any explanation. The model includes a huge amount of information from different educational theories and practice but does not provide any guideline to follow. According to Carvalho et al. (2015), the GOM II model does not show the progression in the relationship between game elements and learning elements over time, and the model's diagram is considered complex and difficult to understand (Arnab et al., 2015).

Another framework is the “I's” framework (Annetta, 2010), which has the purpose of designing educational games and it is built on learning theories and instructional technology. The framework includes six nested elements: identity, immersion, interactivity, increasing complexity, informed teaching, and instructional. Based on the literature, these elements are seen as the most important components to design SGs. However, the framework omits connections to learning theories, and does not explain the interrelations between the nested elements. For instance, the framework discusses the value of identity or immersion in video games in general, but without discussing the pedagogical view of these elements in educational games. Furthermore, the framework does not show how identity relates to informed teaching or interactivity, for example. In-game assessment with feedback is included in the informed teaching component and is mentioned as a promising future area of research.

The Learning Mechanics-Game Mechanics (LM-GM) model (Arnab et al., 2015) aims to transform learning goals and practice into elements of gameplay by providing pre-defined lists of game elements and educational elements. The designer is supposed to draw a type of map by selecting pedagogical elements and fun elements along with their interrelations. The model is limited in considering and discussing in-game assessment and feedback. Also, Carvalho, and Bellotti et al. (2015) argue that LM-GM fails to show the connection between game mechanics and high-level learning objectives (Arnab et al., 2015). In my opinion, this framework is difficult to follow and apply to design GBL.

The Activity Theory-based Model for Serious Games (ATMSG) (Carvalho et al., 2015) adapts concepts from the LM-GM and GOM frameworks to analyse SGs and investigate the connections between learning elements and game elements. Also, it can be used to design SGs. The model links learning theories with game components based on activity theory to identify three types of activities: gaming activity, learning activity, and instructional activity. Each activity consists of a motive, subject, and tool, which link to a sequence of actions over time representing SG components such as tokens, tips, and characters. The model does not consider social structures, especially collaboration and cooperation issues.

Nadolny et al. (2020) displayed two frameworks to select and design GBL. The primary framework includes (immersion, interaction, learner control, assessment, narrative, and interaction) characteristics and the secondary framework consists of (penalty, reward, digital immersion, collaboration, competition, feedback, game choice, sensory element, and support) characteristics. The two frameworks should be combined together, however, the guidance to do that was not provided. Besides, learning theories are not included in any of the two frameworks, though some cognitive tasks were identified as outcomes of specific characteristics without explaining how to facilitate those characteristics to get the expected cognitive outcomes.

The following frameworks are focusing on designing VFTs and some of their concepts such as assessment or specific subjects. These frameworks are recent as the fields of VFTs and combining VFT to GBL is a more recent field of research. There are a good number of studies presenting different technologies to create VFTs however, few frameworks regarding VFTs are presented in the literature.

Pham et al. (2018) present a conceptual and theoretical framework to design VFTs for a specific subject - Mobile Construction Safety. Also, Molan, Weber, and Kor (2022) provide a VFT with an embedded assessment that is designed based on a design template for problem-based VLE. The frameworks that focus on a specific subject or particular learning aspect are less important to be discussed in detail which is the opposite of the proposed conceptual framework of this thesis that can be employed to design VFTs regardless to the subject and can support different types of high-order skill.

FRACH (Andreoli et al., 2017) is a design framework for Cultural Heritage SG which could be considered as a VFT to historical places and events. The framework consists of four phases. The primary phase is a foundational step to identify the initial idea and main learning goal. The Conceptual Phase concerns with historical, scenes, challenges, and storytelling while the development phase involves selecting the game engine and implementing the 3D presentation. The evaluation phase focuses on validation and user experience tests. The framework concentrates mainly on the validity of historical presentation and little attention is paid to game elements with the exclusion of game mechanics. The only learning aspect discussed is defining the learning goal - macro & micro. In my opinion, a framework of Cultural Heritage SG should discuss the balance between the degree of realism and the learning goal and the possibility of overwhelming students with cognitive overload.

Education Through Exploration (Mead et al., 2019) provided design principles of interactive VFTs and has been recently employed to design a number of VFTs (Anbar et al., 2017; Horodyskyj et al., 2018; Mead et al., 2019). The principles are: encouraging learners to explore and discover, crafting a story to guide the VFT, and utilising virtual environments to create learning opportunities that are impossible in actual FTs. All of the three VFTs developed based on Education Through Exploration are multimodal presentation VFTs (see Section 2.3), meaning that they primarily include images, maps, videos, or 360° images and could involve some assessment/feedback during or after the VFT. However, the principles are targeting a specific type of VFTs. There is another example of a framework constructed to support this type of VFTs (multimodal presentation VFTs) - Meaningful IVR learning framework (Mulders, Buchner, & Kerres, 2020). The main purpose of the framework is to guide the use of immersive virtual reality such as VFT in learning environments. The framework works on interweaving the immersive virtual reality into the learning process by mixing and supporting the key features of virtual reality with Instructional design. The framework utilises the cognitive theory of multimedia learning and provides some design recommendations of immersive virtual reality such as concentrating on learning before immersion. As mentioned before in Section 2.3 of this chapter, this type of VFTs (multimodal presentation) usually its strength comes from the supported materials (teacher's instructions or spreadsheet activities).

The Research framework for Immersive Virtual Field Trips (iVFTs) (Klippel, Zhao, Oprean, Wallgrün, & Chang, 2019; Klippel, Zhao, Sajjadi, et al., 2020) is a recent extended framework for assessment and research of immersive learning experiences, particularly immersive VFTs, and consists of two tools. The first tool is a VFT taxonomy which distinguishes three types of iVFT that suppose each to be reflected differently in learning. The first type is a replication of a physical FT; the second type provides visual information that is not accessible in the physical setting; and the third type includes a model or simulation. The second tool is a sensing-scalability trade-off continuum that indicates the sensing capacities, associated costs, and interaction possibilities of various XR systems and how these facilitate learning. The framework mentions the connections between types of iVFT and the XR system for learning; however, there is no explanation of these connections or guidance in how to apply the framework to assess iVFTs. This framework is included in this review even though its purpose is not to design VFTGs; instead, the framework is supposed to be applied to assess a VFTG. However, the research framework is utilised to create the three different types of VFT (Klippel, Zhao, Jackson, et al., 2019; Klippel, Zhao, Oprean, et al., 2020; Sajjadi et al., 2020a) based on the VFT taxonomy tool.

For example, the research framework is facilitated the process of designing VR serious games (Klippel, Zhao, Sajjadi, et al., 2020). This VR serious game called Critical Zone is an example of the third type of iVFT (Sajjadi et al., 2020a) yet there are no clarifications about the supposed connection to learning; the authors cited applying ELT and embedding a narrative into the learning experience in addition to classifying the activities based on the taxonomy of Bloom (Sajjadi et al., 2020b). In my opinion, the research framework is focusing on the technical part of designing or assessing VFTs regardless to their claims of supporting the experience. The framework does not discuss any pedagogical aspects in connection to the three types of iVFTs. Focusing on the third type as it is representing 3D models and is applied to create VFTG, there is a lacking of discussing of game design aspects else narrative.

Chan et al. (2020) provide a framework for evaluating VFTs as GBL. The framework consists of three phases: pre-game experience, the VFTG, and post-game. However, the study only discusses the VFTG phase which contains three dimensions: knowledge enhancement, attitudinal change, and game platform change. The study does not present the method or any learning theories to design the VFTG else utilising story and role-playing elements.

Table 2.1:Frameworks for designing GBL/VFTG.

Type	Focus on	Examples	Advantages	Limitations
Frameworks for designing SGs/GBL	Some game elements	Activities (Chew, 2017) Narrative (Z. H. Chen et al., 2018). Generating scenarios for training (Luo et al., 2017). Engagement and motivation Alexiou and Schippers (2018) .	More focus on one purpose	No clear guide to help game designers/educators to design GBL. Not comprehensive.
	Specific subject areas	Environmental learning (Fjællingsdal & Klöckner, 2017). Health purposes (Ushaw et al., 2015; Ushaw et al., 2017).		
	A particular genre of digital game	Massively multiplayer online role-playing games (Eseryel et al., 2012)		
	issues related to design	Designing assessment (Abdellatif et al., 2018; Graafland et al., 2014; Mitgutsch & Alvarado, 2012) . Difficulty adjustment of tasks (Shena et al., 2019). Reducing the complexity of SG design Westera et al. (2008).		
	General-purpose frameworks	De Freitas and Jarvis (2006)	It includes some essential elements of learning and game design.	It does not provide an explanation to facilitate the designing process.

Table 2.1: Frameworks for designing GBL/VFTG.

Type	Focus on	Examples	Advantages	Limitations
Frameworks for designing SGs/GBL	General-purpose frameworks	Yusoff et al. (2009)	It covers more game attributes and learning elements.	The connections between the framework's components are unclear. It provides a sequential method of design (input-process-output).
		Experiential gaming model (Kiili, 2005)	It can be considered as a designing model of VFTG. It is based on learning aspects: ELT, Flow theory, and immediate feedback.	Two main elements are not included in this model: social interaction and assessment. few game elements (story, game balance, appearance) are considered.
		Problem-based gaming (PBG) (Kiili, 2007)	It applies problem-based learning (PBL).	It ignores the integration of pedagogical theories into game elements. It cannot stand alone to design GBL but could be part of a richer framework used to support PBL. It does not discuss issues related to individual/multiplayer learning in PBG.
		Game Object Model II (GOM II) (Amory, 2007)	The model includes a huge amount of information from different educational theories and practices.	It does not emphasize the value of assessment and feedback. The lack of any guidelines to follow. The progression in the relationship between game elements and learning elements was not considered.

Table 2.1: Frameworks for designing GBL/VFTG.

Type	Focus on	Examples	Advantages	Limitations
Frameworks for designing SGs/GBL	General-purpose frameworks	"I's" framework (Annetta, 2010)	It covers elements that are seen as the most important components to design SGs.	It omits connections to learning theories and does not explain the interrelations between the nested elements.
		Learning Mechanics-Game Mechanics (LM-GM) model (Arnab et al., 2015)	It includes pedagogical elements and fun elements along with their interrelations.	It is limited in considering and discussing in-game assessment and feedback. It fails to show the connection between game mechanics and high-level learning objectives. It is difficult to follow and apply to design GBL.
		Activity Theory-based Model for Serious Games (ATMSG)	It links learning theories with game components based on activity theory.	The model does not consider social structures, especially collaboration and cooperation issues.
		Nadolny et al. (2020)	It can be utilized to select and design GBL.	Learning theories are not included. There is no guidance.
Frameworks for designing VFTs	a specific subject	Mobile Construction Safety Pham et al. (2018)		Limited to one subject.
		Cultural Heritage SG: FRACH (Andreoli et al., 2017)	It can be considered as a VFT to historical places and events.	It concentrates on the validity of the presentation and little attention is paid to game elements with the exclusion of game mechanics. One learning aspect is considered: the learning goal. Lack of discussion of the balance between the degree of realism and the learning goal and the possibility of overwhelming students with cognitive overload.

Table 2.1: Frameworks for designing GBL/VFTG.

Type	Focus on	Examples	Advantages	Limitations
	A particular learning aspect	Molan et al. (2022)		Limited to problem solving
Frameworks for designing VFTs	General-purpose frameworks	Education Through Exploration (Mead et al., 2019)	It involves important principles: encouraging learners to explore, crafting a story to guide, and creating learning opportunities that are impossible in actual FTs.	It targets only a specific type of VFTs (multimodal presentation)
		Meaningful IVR learning framework (Mulders et al., 2020).	It interweaves immersive virtual reality into the learning process by mixing the key features of virtual reality with Instructional design. It utilises the cognitive theory of multimedia learning.	It supports only multimodal presentation VFTs.
		iVFTs (Klippel, Zhao, Oprean, et al., 2019; Klippel, Zhao, Sajjadi, et al., 2020)	It can be used for the assessment and design of VFT. It provides a VFT taxonomy tool and a sensing-scalability trade-off continuum tool that indicates the sensing capacities, associated costs, and interaction possibilities.	No guidance regarding how to utilise the framework. It focuses on the technical part of designing or assessing VFTs more than the learning experience. It does not discuss any pedagogical aspects. It lacks discussion of game design aspects else narrative.

In summary, the literature shows the need for VFTs in GBL. Considering the importance of VFTGs, which provide more than high-fidelity presentation, there have been attempts and efforts to implement VFTGs. A conceptual framework would facilitate and enhance the quality of GBL design for VFTs. There is a need for a framework to guide the process of design and development. In addition, using game elements in virtual fieldwork to engage and motivate learners has been found to be effective (Ashfield et al., 2010; Getchell et al., 2010). VFTGs should not aim to replace traditional FTs (Jacobson, Militello, & Baveye, 2009; Lenkeit Meezan & Cuffey, 2012; Anoop Patiar, Emily Ma, Sandie Kensbock, & Russell Cox, 2017), but can support and enhance learning and be used when it is difficult to provide actual fieldwork for distance learners or disabled learners.

2.5 Assessment

Assessment and feedback are essential steps of the evaluation process and follow each other in a cycle to ensure the progression of learners' learning. Feedback is provided after any type of assessment. The literature review focuses mainly on applications of assessment/feedback, and methods used to implement assessment and provide feedback. Assessment is the process of evaluating learners' ability to learn knowledge or skills. A simple definition of knowledge acquisition is memorizing facts and understanding, and assessment in this case involves measuring the ability to recall memorized knowledge through simple traditional questions. On the other hand, higher-order skills such as analysing, evaluating, and creating require more sophisticated cognitive abilities than those applied at the level of understanding (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). The assessment in the traditional classroom environment can be used in e-learning, such as informal or formal, formative or summative, and individual focused or group focused. Formative assessment aims to improve learning by providing information to help learners reduce the gap between their current performance and learning objectives.

Electronic assessment (E-assessment) tools support different types of assessment (e.g. formative assessments) and also apply a variety of developed techniques. The literature includes five categories of e-assessment tools, which can be grouped to achieved or developed as displayed in Fig. 2.2. The achieved group includes computer-based tests (CBT) and computer-adaptive tests (CAT). The developing group (automatic scoring, continuous assessment, and game-based) includes some achievements and still needs more innovations. Each e-assessment category is defined and explained below in addition to reviewing the literature. Finally, the assessment in VFTs is considered and discussed.

Computer-Based Tests (CBT): The main benefits of CBT are the automation of grading and immediate feedback of overall test grades where paper-and-pencil tests are converted to electronic versions. CBT consists of a bank of questions/items and a test engine. It is a transformation of paper-and-pencil tests to electronic versions. CBT provides many benefits: objectivity, automatic scoring for pre-determined questions, feedback set up in advance by the teacher, and saving time. Selected response test items that form CBT assessments enable standardisation and objectivity to achieve efficient assessment. Assessment creators seek to eliminate the subjectivity derived from human scoring and at the same time require comparability of exam scores among learners based on standardisation (V. J. Shute, Leighton, Jang, & Chu, 2016).

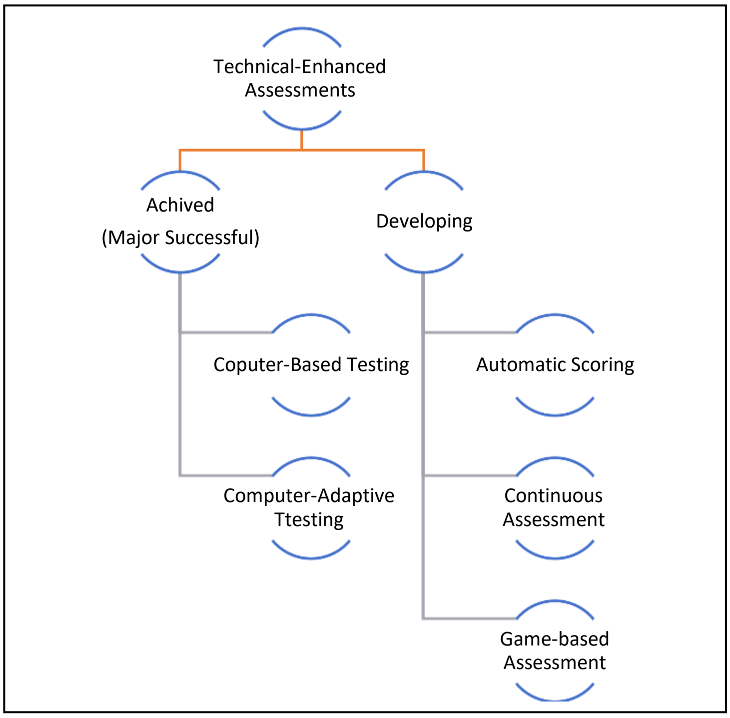


Figure 2.2: E-assessment categories

There are a good number of assessment tools that apply CBT and are available freely (e.g. Hot Potatoes, Quiz Center) or commercially (e.g. Questionmark Perception) and both types are used widely by educational institutions around the world. Also, CBT assessment tools can be stand-alone (e.g. Flashform Rapid eLearning Studio) or part of a LMS (e.g. Model Quiz tool). A number of examples of CBT tools will be analysed and compared below to provide a detailed overview of their features and limitations while a summary is displayed in Table 2.2.

Hot Potatoes (Arneil, Holmes, & Street, 2001) is one of the most used non-profit tools that enables teachers to create question items with pre-determined answers. The types of questions are: JBC (multiple-choice), JQuiz (short-text entry), JMix (arrange words to form phrases), JCrossfill (complete a crossword puzzle), JCloze (fill-in-the-blank), and JMatch (matching). All types of questions can be programmed with the correct answers to automatically score learners' responses. To answer these questions, the learner needs to click, drag and drop, or enter text. Feedback has to be entered by the teacher in advance. Hot Potatoes supports the creation of exercises for learning a second language, as it is based on second language acquisition theory. It has very well-known and basic test-tool features. However, there are a number of limitations: it is not a test-design package; the types of question items are traditional and cannot measure high-cognitive skills; and the opportunities for interaction are limited. In fact, this tool does not provide any protection from cheating or security, nor does it keep records of learners' responses. However, Hot Potatoes is suitable for formative self-assessment (T. Miller, 2009).

Desire2Learn (D2L) Quiz tool is part of LMS that provides different tools to support learning and assessment. The D2L Quiz tool allows teachers to create question items and save them in the Questions Bank. Question items are traditional types: multiple-choice, true or false, fill in the blanks, long answer, matching, ordering, and maths questions. All types can be programmed with correct answers unless they are long-answer questions. The D2L Quiz tool supports different types of multimedia content with a low level of user interaction. The feedback can be pre-scripted by teachers and can be given as overall feedback to individual learners. The D2L Quiz tool overcomes most of the limitations of Hot Potatoes by managing and authoring assessments and keeping records of responses for further tracking and analysing. However, it still has limitations, such as there being no mention of the learning theory that the D2L Quiz tool is built upon, and as with Hot Potatoes, it does not take advantage of multimedia interaction.

Devices such as smartphones, iPads, tablets, and video game consoles can benefit learning and assessment, enriching learners' experience. A number of studies have examined the potential of handheld devices in assessments (Chao hsiu Chen, 2010; de-Marcos et al., 2010; Karadeniz, 2009). One example was created by Hwang and Chang (2011) to implement formative assessment in a mobile learning environment. The main advantage is that it gave learners the experience of learning and being assessed in a real-world scenario without limitations of space and time. Learners had to explore the Chin-An temple in southern Taiwan as a learning activity and were provided with a wireless network connection and mobile devices (PDAs). The formative assessment guided learners to observe and answer test questions with feedback given as hints to direct learners to observe the real-

world environment. However, the types of question items were not clear, and the system did not provide scores or even correct answers in order to encourage learners to find the answers by themselves (Hwang & Chang, 2011). This method promotes challenging and motivated learning. On the other hand, this tool is limited in measuring higher-level thinking skills and the feedback lacks important information. However, the majority of mobile-based assessments allow for the evaluation to occur anywhere and anytime.

QuesTInSitu (Santos, PéRez-SanagustíN, HernáNdez-Leo, & Blat, 2011) is an assessment tool with GPS for use in mobile learning environments. QuesTInSitu uses traditional types of questions: multiple choice, Yes/No, and multiple responses. The questions are linked to geographical coordinates, and the learner has to find the right position to display the test items to engage learners and encourage them to observe the learning. Google Maps is used as a web map application to allow teachers to create a route. The tool provides visual feedback with a green marker displayed on the map when questions are answered correctly, and red otherwise. In addition to visual feedback, the tool provides pre-designed text feedback from the teacher to show the correct answer and the next step. The main advantage of QuesTInSitu is that it enables the creation of authentic assessment experiences that engage learners and ensure interactions with learning environments to assess higher-order skills. However, the limitations are the lack of communication and collaboration in solving test items and the limited types of questions.

Unfortunately, the vast majority of CBT assessment tools transform paper-and-pencil tests into electronic versions using traditional types of selected response question items, which miss two important aspects: measuring high-level cognitive skills and using the power of information and communications technologies (ICTs) (Conole & Warburton, 2005). Despite all of the benefits of the CBT method, its efficiency is considered unsatisfactory, because the same pre-defined set of test items is given to all learners without any consideration of the ability level of individual learners. However, CBT assessments are suitable for formative low-stakes tests.

Computer-Adaptive Tests (CAT): Usually, a fixed set of test items is presented to all learners in traditional CBTs without taking into consideration their abilities. Instead, test items are carefully chosen to cover a wide range of ability levels, from low to advanced. This type can be improved to adapt the questions in the assessment to each learner's individual ability. A CAT would start with a random question at the average ability level, and the remainder of the test's questions would be dynamically selected based on each individual's performance in the test. If the learner answers the question correctly, his/her ability estimation will be increased, which results in presenting more challenging questions. In contrast, if the answer is incorrect, the estimated ability is decreased along with the difficulty of the next question (Lilley, Barker, & Britton, 2004; Weiss & Kingsbury, 1984).

The main benefit of CAT is its efficiency: fewer question items need to be given to learners to measure an acceptable level of accuracy statistically (Linacre, 2000). One of the first attempts at CAT, Weiss (1974) applied a simple method by giving a specific next item after the previous item was answered correctly and giving a different item after a wrong answer was given.

Research has found that CAT is a reliable and valid method of testing that is equal to or even better than CBT. In particular, test length can be reduced by up to 50% with CAT compared to CBT. The main elements of the CAT method are as follows: an item bank that contains a collection of test items from different levels of proficiency to form a full range (latent variable); a termination criterion, which consists of rules to end the test; and item selection, which is an algorithm used to select an item from the item bank to present to the learner. Item response theory (IRT) is a psychometric theory describes the relationship between the learner and the item in a mathematical model (Hambleton, Swaminathan, & Rogers, 1991), and is most commonly used with CAT. IRT mathematically estimates the probability of learners answering a specific test item correctly in order to select the next item. IRT can be combined with different item selection algorithms, stopping rules, and scoring methods. The estimation process can use different numbers of parameters: one-parameter logistic (1-PL) model that requires information on the difficulty level of the items and uses fixed values for item discrimination (amount of information) and guessing (probability of guessing the correct answer); two-parameter logistic model with a fixed guessing parameter; and a three-parameter logistics (3-PL) model that is supposed to estimate ability level faster.

Lilley, Barker, and Britton (2007) created a computer-adaptive assessment of English language and grammar. The adaptive algorithm is based on the 3-PL of IRT and the test process consists of displaying a question, then evaluating the question response as correct or incorrect, then calculating the ability estimation using the 3-PL function, and then selecting the next question based on all previously estimated abilities. This process is reiterated until a specific number of test items has been administered or a maximum time has elapsed.

The experts' opinion about this computer-based test is that it is easy to use and learn and efficient in grading a large volume of learners by reducing error and workload. Also, learners who used the test were fully aware of the adaptive concept from previous assessments such as GMAT and TOEFL and perceived it as a reasonable and fair method of scoring where questions are weighted based on difficulty. The only concern put forward by learners was that they cannot go back to review or change the previous questions during the exam. The study showed that there was no significant difference in performance between the CBT version of the exam and the actual CAT test (Lilley et al., 2004). What distinguishes this assessment tool is the adaptive characteristics of the test algorithm, and the major disadvantage is that no feedback is delivered to learners in any form and there is no mention of adding a feedback mechanism in future work.

The second example from the literature implemented CAT in a different environment – a mobile environment called Computerized Adaptive Test on Mobile Devices (CAT-MD) (Triantafillou, Georgiadou, & Economides, 2008). The test algorithm uses IRT and applies a dichotomous model where questions can be answered with either of two response options: correct/incorrect. This CAT mobile tool is differentiated from the first CAT tool in Table 2.1 by using 1-PL model, which concerns the question difficulty and is the simplest model of IRT. The teachers preferred the mobile nature of the assessment as it reduced the need for computer labs. Also, learners stated that it was an enjoyable and attractive mode of assessment due to the use of mobile devices, and at the same time appreciated the adaptation of the test questions to their level of knowledge; in addition, immediate feedback was provided by notifying learners after each response with a correct check mark or incorrect mark (Triantafillou et al., 2008). In summary, CAT-MD demonstrates efficiency by requiring fewer items to measure learners' abilities. However, it has minimal advantages over a regular CAT in a computer-based or web-based environment because the special features of mobile devices, such as GPS, were ignored. In addition, the feedback is very limited with the assessment tool providing a correct indicator in the form of visual feedback (green right tick).

The third example from the literature is the SIETTE tool (Conejo et al., 2004), which is part of the web-based learning environment. Questions can contain multimedia content and Java applets, which allows the inclusion of interactive elements in the item presentation and answers. The tool allows tests to be created in a curriculum-based structure where each subject is represented with questions; the selection algorithm ensures content balance in the tests. The tool can provide feedback in the form of knowledge of the correct answer. Two main advantages of SIETTE are: using multimedia content and applets to provide questions that are difficult to provide and evaluate in traditional assessments and teachers being able to specify the selection strategy and stopping rules.

The final example is a flexible platform for web-based assessment that allows instructors to implement CAT or CBT assessment (Oppl, Reisinger, Eckmaier, & Helm, 2017). The main advantage is providing a customizable CAT tool that is flexible in terms of testing strategy and the design of the item pool (item bank). Also, the system allows interactive item visualization by external components to provide item presentation. The study proposes the idea of flexible provision of feedback of different amounts and types of information. However, this idea has not been discussed and neither has the embedding of CAT into the online learning platform.

In summary, CAT is more difficult to implement than CBT because of the need for an adaptive algorithm and a larger item bank. Learners usually show concern about not being able to review and modify their answers in CAT assessments. CAT is most commonly used with large-scale tests and provides a minimum amount of feedback.

Automatic Scoring: E-assessment tools that apply different methods and algorithms to score ill-structured responses such as writing and programming tasks. It supports the evaluation of high-order skills in domain-specific tasks such as programming, maths, and science, which require more than multiple-choice questions and correct responses can vary from one learner to another. Researchers from each domain develop methods of assessment to evaluate the features of the domain field and to meet its requirements. A general review of automatic scoring of programming tasks and writing essays is covered in this literature review. Automatic scoring of programming tasks is used for many reasons: the increasing number of learners, saving teachers' time to do other things, and providing real-time feedback. A very basic method of automated programming assessment applies a comparison between the learner's solution and model programs provided by the teacher and measures the similarity (Rahman, Ahmad, & Nordin, 2007). The evaluation process uses static analysis and cannot support dynamic assessment, and the teacher has to provide more than one correct program model to cover many possible variations in learners' solutions. Assessment of programming tasks depends on specifying measurement values/features based on the learning goals and then extracting them.

The most used method of automatic assessment of programming applies test cases, where the learner's solution is run on test cases to assess the compilation and execution (Saikkonen, Malmi, & Korhonen, 2001; Spacco et al., 2006). Then, the learner's result is compared to the result of a model solution provided by the teacher. Feedback provided in this assessment method is mainly limited to knowledge of result, such as identifying the test cases where the learner's program fails – more details are provided in the following section about feedback. Some assessment tools use a visualization mode to represent these two feedback types (Edwards, 2003). However, the feedback does not provide the required information to improve wrong solutions but encourages learners to search for mistakes (Ihantola, Ahoniemi, Karavirta, & Seppälä, 2010).

Automatic assessment of ill-structured responses such as writing essays considers different measurable values of syntactic and semantic features (Valenti, Neri, & Cucchiarelli, 2003). Taking a broad view, automatic essay scoring (AES) relies on data-driven methods starting with selecting features based on statistical approaches to predicting scores from essay training sets. Syntactic features include the essay length, sentence structure (e.g. relative parts of speech, count of prepositions), and differences in word length. An example of an automatic scoring tool that uses statistical approaches based on a hypothesis that the measurable "*proxes*" reflect the quality of wiring is Project Essay Grade (PEG) (Page, 1994). PEG used human-graded essays as training essays and applied regression coefficients to mark learners' essays.

On the other hand, semantic features that can be evaluated involve discourse structure and vocabulary usage. E-rater is another automatic scoring tool that evaluates syntactic and semantic features (Attali & Burstein, 2006). E-rater uses natural language processing (NLP) for parsing the whole essay to extract linguistic features with a combination of statistical techniques. The strengths of E-rater are providing a fast and automatic tool for essay assessment along with proven validity and reliability. However, it is limited to a specific number of topics as trained on human-rated essays of these topics, and knowledge of result (KR) feedback. In general, AES is suitable for high-stakes assessments which provide simple scoring reports without any instructional feedback that would improve learners' performance.

Continuous Assessment: uses educational data to perform formative assessment and tutoring. They aim to provide real-time information about the learner's progress and provide feedback via applying data-mining/learning analytics techniques. It is applied to web-based learning environments, LMSs, or historical educational data. Assessment can be done by tracking learners' activities and mining log files. Educational Data Mining (EDM) is used to evaluate learners' performance to provide feedback, recommended materials, or early help and support. General tasks related to the evaluation are based on capturing and mining learners' usage and behaviour in web-based learning environments to gain insights about learners' performance.

A number of studies have used different EDM techniques to classify, predict, and find patterns in learners' performance and dropout possibilities in order to provide early support to learners who need extra help and feedback to improve their performance and reduce chances of failure (Hwang, Hsiao, & Tseng, 2003; Minaei-Bidgoli & Punch, 2003). F. Yang, Li, and Lau (2013) presented a model for designing learning paths and evaluating learning outcomes which can be used in LMSs. It is a fine-grained outcome-based learning path model that supports teachers to design learning paths by explicitly defining pedagogy and assessing generic skills in addition to subject-specific knowledge. The key element of the learning path is a learning activity that consists of a set of learning tasks, each task should train and evaluate a particular type of student ability. This method allows the reusing of pedagogies of defined learning activities and assessing more than knowledge-based questions.

Baradwaj and Pal (2012) used a decision tree algorithm to predict learners' final grades so that learners can get help and support from teachers. Learning databases with information such as class tests, assignments, and attendance were used from previous years to generate classification rules (IF-THEN) to identify learners who need more attention and support before dropout or failing. This paper presented a promising mode of assessment and left the entire process of how to provide this support (e.g. providing feedback) to individual teachers to handle.

Data mining and text mining can be used to support teachers to assess and provide feedback manually. Dringus and Ellis (2005) applied data mining and text mining to extract data from online discussion forums to provide a more manageable overview of information. Temporal participation indicators were extracted such as time and sequence to show how the discussion progresses over a specific period of time by individual or group. For example, the teacher can inquire into the degree of presence in the discussion forum by specifying a date range, authors, and a number of replies.

In addition, text mining is used to find out whether resources or references were shared. However, combining the results of data mining with a more qualitative evaluation of learners' replies was not included in this study. Data mining and text mining were used to simplify manual assessment and not to provide a full e-assessment or feedback.

Game-Based Assessment: Embedded and stealth assessment in SGs/GBL that evaluates the player's knowledge and skills continually and invisibly. It is a promising method of formative assessment where games combine powerful entertainment features with assessment. GBL is an encouraging environment that mix learning with experiences of real-world scenarios to enhance problem solving, social communities, and conceptual learning, and the literature shows a strong relationship between engagement and learners' achievement, especially at the academic level (Finn & Rock, 1997; Fredricks & Eccles, 2006). The literature has revealed that tasks embedded in gameplay not only increase motivation and effectiveness but also performance (Clark, Nelson, Sengupta, & D'Angelo, 2009; Mekler, Brühlmann, Opwis, & Tuch, 2013; Ninaus, Kiili, McMullen, & Moeller, 2017; Ninaus et al., 2015; J. C. Yang, Chien, & Liu, 2012).

To be considered as a viable educational tool, GBL has to show the significance of learning progress and outcomes, especially for learners. To assess learners in games, some questions arise: how to know that learners have learned what they were supposed to learn; which data indicates the acquisition of which conceptual knowledge or cognitive skills; and how to track learning progress for individuals and teams. Assessment in the game can be of two types: 1) overall assessment at the end of game or level (summative), and 2) continuous assessment during the game playing (formative).

Many studies in the literature have discussed the option of assessment in games with the tracking of players' behaviours based on the amount of collected data derived from interaction with learning environments (S. Chen & Michael, 2005; dos Santos Nunes, Roque, & dos Santos Nunes, 2016). Unfortunately, few studies have proposed methods of actual evaluation in games by using tracking and players' interaction data. The implementation of game-based assessments can be distinguished into three categories: external assessment, game scoring, and embedded assessment (Ifenthaler, Eseryel, & Ge, 2012). Each type is discussed and analysed in the following selected examples.

External assessment is applied by providing an assessment that is not part of the game environment. Learning progress/outcome is measured after or during gameplay by giving learners traditional test questions such as multiple-choice or essay questions to assess what they learned from the game. Bloomfield and Livingstone (2009) added e-assessment to the virtual world *Second Life* in the usual and known way. A tool called quizHUD was used to display a web window during *Second Life* and presented test questions and at the end of each test, a final grade is displayed. The aim is to provide an objective assessment as part of the virtual experience in the same place as the learning took place (Bloomfield & Livingstone, 2009). This method of assessment in the VLE could be considered as a first attempt, but technologies today can support more advanced methods of assessment. However, the obvious limitation is providing qualitative feedback. Learners do not obtain any information about correct and incorrect answers and have no guidance to improve learning. In fact, this type of assessment disturbs the learning process. Bellotti, Kapralos, Lee, Moreno-Ger, and Berta (2013) present another example that is used the same mechanism of assessment (multiple-choice questions and clicking on objects) but with different domains and different gameplay. This type of assessment is utilised for a different purpose - avoiding test anxiety or motivating them to practise via formative assessments (Tsai, Tsai, & Lin, 2015). The well-known game tic-tac-toe, is a turn-based game strategy, designed to assess 9th grade learners in energy education using traditional multiple-choice questions attached to the gameplay. To place a piece on one of the game grids, a player should answer a question correctly, and immediate elaborated feedback is provided each time in the form of a text message. Unfortunately, there is no explicit learning in the gameplay with this type of game-based assessment except for learning from the formative feedback. This assessment did not use advanced technology to evaluate learners while playing the video game.

(T. H. Wang, 2008) argues that providing an electronic version of paper-and-pencil tests will not impact learning or assessment effectiveness, and game-based assessment has to apply strategies to enhance feedback provision and the human–computer interactions. To do so, (T. H. Wang, 2008) is implemented with different strategies such as “repeat the test”, “prune strategy”: which remove one incorrect option, and “call-in-strategy”: the rate at which other learners have chosen a particular answer. The author assumed these strategies create gameplay effects in tests and proved using an experiment that this has a significant effect on learning. Also, hints and references to find the correct answers are provided. However, this assessment provides a traditional type of question (multiple-choice questions in particular), and in my opinion this cannot be considered as a quiz-game-like assessment because it does not provide rules for play or a storyline, and also nothing about these strategies would motivate or challenge learners compared to regular video games or even games such as card games or board games.

Another recent example involves transforming paper-and-pencil tests to game-based assessments to screen reading difficulties assessment (Hautala et al., 2020). The study applied some learning findings to create the assessment tasks and employed statistical analysis to prove that the game-based assessment is acceptably reliable. The assessment provides immediate feedback which is not available in the paper-and-pencil version; however, the game-based version seems to be missing the game elements/mechanics.

Game scoring in SG/GBL is used as an indication of gains in knowledge/skill and stars/badges are awarded based on this while hints to complete tasks can be viewed as a feedback. Games focus on achieving targets, the time in which quests are completed, and overcoming obstacles to increase players' scores. Losing a battle or resources in games reduces a player's number of lives, and solving a major task allows players to level up. This method is considered as a method of assessing learning progress/outcomes in GBL (Kelly et al., 2007; Ninaus et al., 2017; Pasquier et al., 2016). Game scoring is a very simple mode of assessment that ignores a huge and detailed amount of data collected from players' interaction with the learning context.

Immune Attack (Kelly et al., 2007) is a SG that teaches immunology with a 3D representation of biological structure and functions. The game depends on motivating learners through levels of challenges and requires an increasing gain of knowledge to succeed in completing tasks. The game does not assess the player directly but requires the achievement of learning (game scoring) to progress in the game. Feedback provides two types of information: hints to complete tasks while playing, or information to learn from in the case of failing a task.

The French Health Service produced a SG to train soldiers on how to manage combat casualties and in which order of priority to save lives (Pasquier et al., 2016). The game simulates a 3D hostile medical environment with specific tasks to manage, and the soldier receives scores and feedback. The game is called Medusims and is a first-player shooter genre for a single player designed to enhance cognitive training and increase procedural skills. The scoring system rewards soldiers with gold, silver, or bronze medals. The feedback can be motion visual(an instructor avatar shows a procedure) or overall text feedback. However, this paper did not provide enough details to understand the evaluation and feedback mechanism.

Another example designed with game scoring is used to measure 11-year-old learners' knowledge of fractions (Ninaus et al., 2017). This tool is called Semideus Exam and does not apply the traditional question items format. One of the tasks involves estimating the position of a fraction on a number line. The learner avatar has to solve this task while playing the game by locating a gold coin based on its value. Tasks are part of the gameplay in a motivated and non-invasive way. The scoring system relies completely on the game scoring mechanism. Each inaccurate answer results in a loss of virtual energy and accurate answers are awarded with coins. In addition, the provision of feedback is immediate in different modes (motion, colour, etc..). The assessment is the game tasks themselves and does not disrupt engagement with the gameplay. This method of assessment does not evaluate the hidden learning behaviours of the learners playing (ignoring learning analytics) and just measures obvious knowledge that can be measured with traditional CBT assessment. It does not mention any adaptive mechanism for the tasks.

Embedded assessment (stealth) is a performance-based assessment which provides an invisible assessment that is merged with the fabric of the game learning environment without interrupting the flow of the game or engagement in the learning process. Embedded assessment observes and tracks learners' behaviours while they play to form inferences about their knowledge and skills, and the assessment results can be used to provide adaptations and feedback.

The first example of embedded assessment to be discussed aims to assess higher-level skills by analysing physiological data and players' gameplay behaviours. One of the activities in the Crystal Island game encourages the player to read books or articles (Taub et al., 2017) to solve game tasks. Players have to decide if they need to increase their knowledge by more reading; which involves scientific reasoning and self-regulated learning. The player's decision infers a metacognitive judgement and starting to solve the tasks reflects a cognitive learning strategy. The assessment methodology applies log data to measure clicking on specific book titles, and eye tracking provides evidence of fixation on areas of interest in reading. variables of assessment are either extracted from log files or calculated. In general, limitations are that the learners' data are not used to adapt the game activities based on their abilities, and the assessment ignored collaborative tasks. Also, sequence mining is used to capture patterns of learners' behaviours and to classify learners into groups based on their level of skill.

Another SG that provides an embedded assessment tool is Grand Theft Auto IV (GTA IV) (Zielke et al., 2009). The GTA IV game is a virtual cultural trainer designed to increase players' cultural knowledge. The assessment is internal with feedback given in the form of a score (KR) and possible choices of interactions (KCR). The assessment method is based on analysing logs of interactions between the player and NPCs in addition to textual analysis of their conversation. However, no further information is provided on the evaluation mechanism. Shute and a number of his colleagues (V. Shute, Ke, & Wang, 2017; V. J. Shute, 2011; V. J. Shute & Ke, 2012; V. J. Shute et al., 2009) proposed and implemented an invisible method of assessment – stealth assessment – that will not interrupt the learning process or disrupt the flow by using evidence-centred design (ECD) and Bayesian network analysis to monitor the learning that occurs in the game learning environment and to elicit evidence about the learner.

Engage (Min et al., 2019) is a more developed stealth assessment in GBL where deep learning is applied to formulate evidence models automatically in ECD instead of employing statistical rules created by domain experts and then programmed utilising Bayesian networks.

Embedded assessment is invisible in game learning environments, which benefits the learning process by ensuring continuous learning without interruption and eliminates the exam anxiety usually caused by traditional assessment. On the other hand, stealth assessment is still in the development stage and needs further improvements in order to provide feedback, adaptive tasks, and collaborative task assessment.

<u>VFTs Assessment:</u> the assessment associated with FBL and is usually performed after completing the FT by a day or even a week which reflects the post- phase of the FBL structure that is explained in Section 2.2.1. Some of the reasons for delaying the assessment after completing a FT are the limited time and the difficulty of assessing a large number of students by one or two teachers in the field. The assessment tasks include discussion, presentation, or writing a report. However, a popular assessment task that is performed by students during the actual FTs is notetaking or writing a journal. The VFTs in general can benefit from this advanced technology as it creates infinite possibilities for assessing the students in the VFs.

The literature shows that game designers/educators of VFTs adopt a new method of mixing during and after assessment tasks in addition to learning activities such as lab assignments (Klippel, Zhao, Jackson, et al., 2019; Klippel, Zhao, Oprean, et al., 2020;

Marshall & Higley, 2021), notebook/making a geological map (de Paz-Álvarez, Blenkinsop, Buchs, Gibbons, & Cherns, 2021), and report/discussion in a chatroom (Pham et al., 2018). However, few VFTs provide only activities such as sketching (Dolphin, Dutchak, Karchewski, & Cooper, 2019), and only observation without any activities (Cheng, 2021; Cheng & Tsai, 2019; Han, 2020). However, assessment tasks can be applied only after completing the VFT such as report and presentation (Seifan et al., 2020). All of the previous examples of assessment are performed and evaluated outside the VLEs either on paper or electronically even if the process of measuring something or collecting data has been performed inside the VFT.

Few examples were found regarding embedded assessments in VFTs. For instance, the immersive virtual environment of bushfire safety (Molan et al., 2022) included embedded assessment (identifying flammable items and suitable places to shelter). Another example is the mobile construction safety education VFT (Pham et al., 2018) as the embedded assessment consists of multiple-choice questions after investigating construction hazards and writing a report where it should be written and submitted inside the learning environment.

2.6 Feedback

Hattie and Timperley (2007) define feedback as "information provided by an agent (e.g. teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding". Feedback consists of two pieces of information: measuring the correctness of learners' work and the gap between actual performance and desired-level performance (Ramaprasad, 1983; Sadler, 1989), and it is considered as pivotal to enhancing learning (V. J. Shute, 2008). Also, feedback has a powerful influence on developing the learning experience and teaching (Brinko, 1993; Hattie & Timperley, 2007). However, a few studies have stated that feedback has no effect or decreases learning (Kluger & DeNisi, 1996; Mory, 2004). The main features of effective feedback are: timing, complexity/length, and functionality (V. J. Shute, 2008).

Feedback differs in the time of delivery, type of provided information, and mode of feedback, as shown in Fig. 2.3 and explained in the following section in detail. Based on timing, feedback can be categorized as: immediate feedback (IF), which is instant feedback delivered to the learner after they answer a question or complete an assessment, and delayed feedback (DF), which is provided after the learner completes the

Table 2.2: Examples of CBT and CAT assessment tools

Type of Assessment	*Author/Tool name*	*Question Items*			*Test Type*			*Feedback Provision*				*Type of learning environment*	*Using learning theories*	*Adaptive*
		Pre-determined	Open-ended	Interactive	Individual	Self-assessment	Peer-assessment	Score	Correct answers	Auto-text	Pre-designed			
CBT	Hot Potatoes	√	√			√		√	√		√	Web-based	√	
	D2L Quiz tool	√	√		√	√		√	√		√	Web-based		
	Hwang & Chang (2011)	√									√	Mobile-based		
	QuesTInSitu	√		√				√	√		√	Mobile-based		
CAT	Computer-adaptive Test for English Language (Lilley et al., 2004)	√			√							Computer-based		√
	CAT-MD	√			√				√			Mobile-based		√
	SIETTE			√	√				√			Web-based		√
	Flexible CAT	√	√	√	√							Web-based		√

assessment by a certain time. The level of information provided in the feedback message can be of three types:

- Knowledge of results (KR): tells learners if the answers are correct or incorrect without giving the correct answers.
- Knowledge of correct response (KCR): provides the correct answers.
- Elaborated feedback (EF): provides extra information or hints related to questions, or guides learners to reveal the correct answers.

In addition, elaborated feedback (EF) can be explicitly given to learners, which includes direct instructions to solve tasks, or in an implicit way that can give hints and let learners explore in order to reach the correct answers.

Feedback messages can be delivered in different modes: oral, written, and visual (Brookhart, 2008). Each different feedback mode has its own distinguishing advantage. All feedback modes can be delivered easily in e-learning and have the benefit of real-time delivery.

With the dramatic development and improvement in technologies, feedback can be given to learners in different modalities: visual, auditory, haptic, and even multimodal form. Visual feedback can take different formats such as text, which is the most commonly used, and a graphic representation with static images or even with animation could help to deliver the feedback message easily and clearly, such as in training skills (Pasquier et al., 2016) or motor learning (Sigrist, Rauter, Riener, & Wolf, 2013), or when trying to get young learners' attention and motivation (Ninaus et al., 2017).

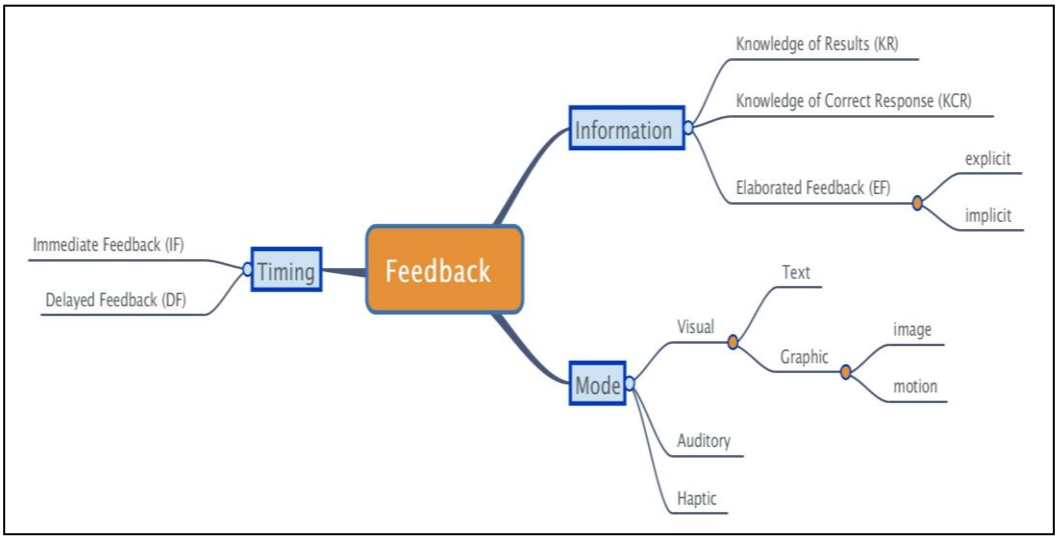


Figure 2.3: Feedback categories.

Research shows a strong relationship between the timing of feedback and enhancing the learning outcome (Annett, 1969; Espasa & Meneses, 2010; V. J. Shute, 2008). Therefore, immediate delivery of feedback helps the learner to focus and improves their knowledge and skills (Denton, Madden, Roberts, & Rowe, 2008; Jordan & Mitchell, 2009; Narciss & Huth, 2006). The power of e-assessment in general, regardless of testing methods, relies on immediate feedback as in real-time and automated feedback messages. For example, Wu, Hwang, Milrad, Ke, and Huang (2012) found that immediate feedback with real-time assessment of complex tasks, which involved creating concept maps for medical and surgical nursing knowledge, significantly improved learning achievement.

Feedback can be provided to learners mainly 1) by instructors, 2) by peers, or 3) as automatic generated feedback. Instructors and peers nowadays use technology as a medium through which human feedback is given. Teachers used to provide feedback to their learners during classes or wrote feedback on assignment papers. However, technology plays the role of delivering feedback electronically via e-mail, or via messages/audio in LMSs (Nicol, 2007). In addition, peer feedback has been proven to be effective (Schultz, 2000), and also can be provided through communication technologies such as discussion forums (Van der Pol, Van den Berg, Admiraal, & Simons, 2008). On the other hand, automated feedback generated by the software is considered effective and cost/time saving (Ware & Warschauer, 2006). Automated feedback relies somewhat on e-assessment and varies from a general report after completing an assessment to step-by-step feedback while solving the assessment. Feedback-generating mechanisms can be categorized into different groups:

Pre-defined Feedback: Question items with pre-defined answers that are usually given to learners in CBT and CAT provide pre-defined feedback to each response. Feedback in this case includes KR and KCR and can be provided immediately after answering a question or after finishing the whole assessment in the form of a report. EF could be provided during the assessment upon the learner's request to find the correct answer (Z. Wang, Gong, Xu, & Hu, 2019) or as a part of a test report to improve the learner's learning, such as pointing to revision topics.

Model Feedback: Specific models can be created to provide feedback, such as an error model that requires a reference solution and a list of errors and create rules associate each error with potential corrections as feedback (Singh, Gulwani, & Solar-Lezama, 2013), or if-then model feedback in a tutoring system (Roll, Aleven, McLaren, & Koedinger, 2011). The if-then model feedback-generating mechanism is used with problem-solving tasks to reach answers in steps. If a learner takes a specific step, then he or she will receive certain feedback based on that step. For example, if a learner responds incorrectly, then a feedback message is given with more explanation about the question or instrumental information to enable them to reach the correct answer. This feedback mechanism is often applied in e-tutoring systems (Roll et al., 2011).

Data-Driven Feedback: Historical data are used to provide feedback in different domain-specific tasks such as programming. In addition, the data-driven method is applied to provide general feedback in learning environments (e.g. web-based courses, LMSs). However, it is limited to giving recommended materials in real time or a delayed general report of progress specifically for teachers rather than learners. On the other hand, data-driven feedback for domain-specific tasks can provide more information to help learners improve their learning. In general, data-driven feedback applies machine learning/data-mining techniques to extract features from learner data and then provide appropriate feedback sentences (e.g. information to correct mistakes, guiding learners to the next step, and example of the correct solution). In programming, historical learner data are used to generate next-step feedback. For example, the Hint Factory algorithm uses a Markov Decision Process (MDP), which is applied to logic proofs and programming to provide feedback by using historical learners' data to find the next step from a partial solution (Piech, Sahami, Huang, & Guibas, 2015; Price, Dong, & Barnes, 2016). This method does not require reference solutions from teachers or test cases that are limited by reliance on one assignment but requires a large amount of previous data collected for a specific task. NLP and the statistical machine learning method are applied to different data to provide quality feedback for essays. Detecting errors in English grammar requires corpus-based and statistical approaches to train on a large volume of edited sources (newspapers). However, applying data-generated feedback to writing essays cannot provide feedback regarding content or logic.

Continuous Feedback: Feedback in GBL should be continuous during the learning process and assessment tasks. It is promising for improving learning and supporting learner progress, but still needs more research. Current GBL provides feedback based on a game feedback mechanism (e.g. rewards, points, and avatar progress) (V. Shute et al., 2017), which fits the reinforcement/punishment feedback mechanism.

2.7 Geographical Video Games

Several video games teach some geographical topics such as *VR-Engage*, and *Global Village*, in addition to a mix of simulation or VFTs with games. Also, some entertainment video games containing knowledge could be employed to teach players about geographical topics such as volcanoes. A brief review has been conducted to show some examples of related work to the research prototype, Island of Volcanoes, which is a VFTG designed and developed based on the proposed framework as a method of evaluation starting with video games that are developed with teaching and learning purposes and then covering entertainment video games.

SimCity is a simulation game for practising urban geography with four updated versions (Gaber, 2007). Players need to observe, analyse, and plan to simulate real planning problems. The game aims to let players learn about the multi-dimensional systems of planning cities, gain problem-solving skills, and demonstrate creativity in planning. The learner plays the role of a planner of a virtual city and manipulates variables such as land use, the water system, fire safety, parks, and education. Players analyse data and information as they make decisions. SimCity is an excellent simulation of urban planning with layers of models for different variables such as traffic, projections, and economic growth. However, SimCity is a simulation and is not a SG or VFTG; in other words, it is not built based on learning theories or pedagogical aspects of FBL. There is no mechanism of assessment or feedback provision because there are no extrinsic goals except for self-defined goals; if these were included, that could lead to the game being more engaging. As a result, the simulation could easily fail its educational purposes and not meet its learning objectives without instructors planning to integrate it into a bigger learning structure, such as preparing learners before starting the simulations, guiding the learning during simulation, and assessing learning outcomes via writing or by looking at learners' work inside the simulation.

VR-Engage is a virtual reality game for practising factual knowledge of geography and reasoning ability (Virvou, Katsionis, & Manos, 2005; Virvou, Manos, Katsionis, & Tourtoglou, 2002). The player has to answer questions about geography to open closed doors and obtain points. Human Plausible Reasoning theory is a cognitive theory to stimulate reasoning and is applied to help players negotiate with a dragon to guess the correct answer. The game mainly tests players on factual knowledge and provides a form of learner–teacher communication through interaction with non-player characters. In addition, the game applies some game elements to support learning such as story, interaction, and feedback. However, the framework did not play a role in designing the game.

Global Village is a virtual world developed as a part of Quest Atlantis, which is an educational game (Tüzün, Yılmaz-Soylu, Karakuş, İnal, & Kızılkaya, 2009). The storyline is about helping some children to go back to their home countries after attending a festival in Turkey. The avatars represent players, and there are 3D implementations for non-player characters of lost children, flags, and artefacts. The game applies three learning theories: experiential, inquiry-based, and collaborative learning and two main tools: Q-mail and a virtual backpack. The game design applies learning theories without following a framework; however, the connection between learning theories and game elements is not clear.

Where in the World is Carmen Sandiego? is a role-playing game that teaches factual knowledge on geographical locations (Charsky, 2010). The player becomes a detective trying to catch a thief, Carmen Sandiego. It is a typical edutainment game that creates a fictional context and uses role playing to teach the identification of geographical locations, which is a lower-order thinking skill. The game design did not follow learning theories or a framework.

St. Vincent's Volcano (Mani, Cole, & Stewart, 2016) aims to educate secondary school learners about historical eruptions and possible future volcanic hazards. It is a SG designed based on ELT where the player can watch visualizations (historical events and hazards) and the interactions are limited to clicking icons to display information. The game design ignored game mechanics and elements; the only game element the designer applied was the rewording of multiple-choice questions. As a result of not following a design framework, the connection between learning theories and the few game elements used in the game is misimplemented. For instance, the reflective stage of the ELT is linked to clicking icons to display information about volcanic hazards.

Disaster Risk Reduction for Earthquake mobile application (Winarni & Purwandari, 2018) is designed and developed to educate and prepare elementary school learners for dealing with earthquake events. The application includes two scenes: an educational game and an earthquake simulation. The design of the educational game did not involve learning theories, a framework, or game mechanics/elements. In fact, the game is more of an interactive video guiding the learner on what to do before, during, and after an earthquake than a game.

The entertainment video games generally display mixed authenticity levels of volcanic visual representations (McGowan & Scarlett, 2021; "Monster Hunter Generations: Ultimate," 2018). The representation can be overstated such as the size of stratovolcanoes ("LEGO Marvel Super Heroes 2," 2017; "Monster Hunter Generations: Ultimate," 2018; "The Shadow of the Tomb Raider ", 2018), while the flow of lava usually represented accurately ("LEGO DC Supervillains," 2018; "Monster Hunter Generations: Ultimate," 2018; "Subnautica," 2018). Entertainment video games have the potential to teach players about volcanoes and their hazards (McGowan & Scarlett, 2021) in a risk-free, motivated, and attractive learning environment. However, GBLs have been designed with the purpose of teaching, providing accurate factual knowledge, and connecting the required learning theories to game mechanics/elements. On the other hand, entertainment video games are designed mainly for fun and may provide implicit learning. Entertainment video games can be used in education with the guidance of educators to assure the accuracy of volcanic visual and hazard representations. A standing example of entertainment video games that are utilised for teaching and learning is Minecraft (Hobbs, Stevens, Hartley, & Hartley, 2019).

Minecraft is an open-world video game where players solve problems by buildings blocks. The game is employed to teach learners about environmental science such as volcanoes. The game provides a learning experience that is similar to FBL but in a very basic way and with limited representation and knowledge. However, teachers (Hobbs et al., 2019) used the game to attract children by applying some learning theories to ensure they achieved learning goals and objectives. In other words, teachers needed to contextualize the learning process by providing an introduction about the topic and hands-on experiment before playing, and also discussing some concepts during the game. Regardless of the excellent game mechanics, Minecraft does not include FT components or learning elements, which means that good game mechanics and elements are not connected to learning and cannot be effective without the teachers contextualizing the learning process.

This chapter highlighted some of the previous work of technology- associated field trips and frameworks for designing SG/GBL. The review shows the necessity of VFTGs as an alternative solution when FTs are inaccessible which leads to the need for a conceptual and theoretical framework to design VFTGs. This thesis is proposing a conceptual and theoretical framework to design VFTGs. The following chapter is explaining the research methodology and the experiment design that was implemented in this thesis.

2.8 Summary

The first question (***What is the current intersection between GBL and FBL?***) of this thesis is answered and summarized in the following points:

- There are some differences between FBL and GBL:
  - The characteristics of learning environment settings are more important in FBL because the learning environment represents a fundamental source of learning content.
  - FBL and GBL require learning theories and pedagogical aspects that ensure effective learning and enhanced outcomes. Each mode of learning may employ different learning theories. However, ELT is a well-established theory utilised in both modes.
  - FBL activities are manipulated by: time/location, and participation/autonomy (Kent et al., 1997). FBL activities are built upon different locations and different times when considering historical locations, which are assumed to provide the necessary learning content and context.
- GBL contributes to FBL in different ways:
  - Mobile learning games are utilised during physical FTs to help learners by guiding them, encouraging exploring and engagement, or collecting data.
  - VFTs are implemented in game engines such as Unity 3D and Unreal to achieve just a high-fidelity virtual presentation of a physical location without any game elements or learning activities.
  - Few VFTs have been designed and implemented to provide GBL, yet they are more like gamified VFTs than a GBL.
- Literature is rich with different methods of VFTs such as: web-access (Anoop Patiar, Emily Ma, Sandra Kensbock, & Russell Cox, 2017; Seifan et al., 2020), Virtual reality (Cheng & Tsai, 2019; Han, 2020), or remote-access (Palaigeorgiou et al., 2017; Stephens et al., 2016). However, research on VFTs

as a GBL continues to be very uncommon.

The second question (***What are the possible strategies to connect FBL/GBL aspects to game design aspects?***) is answered partially in this chapter and summarized in the following points and the second part of the answer is provided in Chapter 4 as a conceptual framework:

- One of the similarities between FBL and GBL is that both learning modes utilised learning theories such as ELT which can be employed to develop a connection between them in the process of design.
- FBL has a specific structure (pre-, during, and post-) field trip. Studies reveal that this structure promotes the creation of an effective fieldwork experience and learning outcomes. This structure is not applied to GBL unless it is GBL for FTs.
- High-order skills of FBL have been almost overlooked in the previous examples of VFTG unlike GBL and can be utilised to form a connection.
- E-assessment has different methods such as CBT and CAT. However, assessment in GBL can be implemented as a simple scoring mechanism, stealth mechanism where assessment is embedded within the gameplay, or external assessment (not part of the gameplay or game environment). Developing FBL as GBL can utilise and improve the same methods of assessment to fit the nature of VFTGs.
- Feedback can benefit from the advanced technologies of GBL to be delivered in real-time with different lengths of information (KR, KCR, or EK) and in different modes (visual, auditory, or haptic).
- Assessment and feedback of VFTGs can advance VFTGs being delivered in real-time instead of being delayed as in physical FTs.
- Designing GBL based on a framework influences learning performance and leads to the desired outcomes. However, the literature shows limitations of applying theoretical/conceptual frameworks (Nadolny et al., 2020; Pellas & Mystakidis, 2020). Researchers in general focus on two aspects (game elements and learning aspects) when designing or evaluating GBL. Yet the connection or the link between them may not be considered (Acquah & Katz, 2020; Calvo-Morata, Alonso-Fernández, Freire, Martínez-Ortiz, & Fernández-

Manjón, 2020; Divjak & Tomić, 2011). For that, connecting game elements to FBL aspects (learning theories, high-order skills, and assessment/feedback) would support the design of VFTGs.

The review reveals some limitations of geographical video games that can be avoided by following a conceptual framework to design VFTGs:

- Some geographical video games are not built based on learning theories or pedagogical aspects, or they did not contain a mechanism of assessment or feedback provision.
- The connection between learning theories and game mechanics/elements are missing or not clear.
- Geographical video games mainly teach players factual knowledge about geographical topics.
- Educators need to integrate geographical video games into a bigger learning structure when learning theories are missing.

Chapter 3

3 Research Methodology

3.1 Introduction

This research aims to facilitate the process of designing and developing VFTGs. To achieve this goal, a conceptual framework is constructed which creates three links to connect learning aspects to game aspects via: ELT modelling, game modelling, and world modelling in addition to modelling evaluation and skill in VFTGs.

The research methodology involves discussing the utilised learning and game design aspects to establish the required connections of the conceptual framework. Also, the research methodology includes developing a prototype to evaluate the proposed conceptual framework. In addition, quantitative and qualitative methods are applied to validate the opinions of game designers/educators and to verify the effectiveness of the proposed conceptual framework. Quantitative methods are the key research approaches that performed. The quantitative data is collected to calculate study variables and verify hypotheses. Pre/post-test, questionnaires, and log game data are applied to gather quantitative data. On the contrary, the qualitative method is performed as a subjective assessment of perceptions, and opinions on usability and usefulness. Mainly, open-ended questions are utilised to recognize subjective game designers' views and collect their feedback for improvement.

In this chapter, the main aspects of learning and game design that are employed in creating the proposed conceptual framework are explained. These aspects are ELT, game internal economy, and game machinations. Also, the research design section introduces the study design that followed. While the research elements section presents the main elements of this study: instruments, participants, variables, and data analysis. Study Instruments involve the methods that are applied to collect data based on the research design aiming to prove the effectiveness of this research, and participants are defined and the criteria they selected based on. Though, study variables are specified according to the research experiment in addition to their possible effects. The data analysis is identified as statistical analysis tests that are suitable for the research experiment and study variables. The section of research study ethics shows that this research follows ethical principles through scientific research. Finally, the summary section recaps the research methodology.

3.2 Aspects of the Framework

This section will explain these strategies to connect FBL aspects to game design aspects as a part of answering the second question of the research in chapter 1 (***What are the possible strategies to connect FBL aspects to game aspects?***). Some of these aspects are: experiential learning theory (ELT), the game internal economy, and game machinations. These aspects are explained briefly in the following sections as they are necessary to comprehend the proposed framework in Chapter 4. However, the connection between FBL and game aspects involves more (assessment, feedback, and high-order skills) than these three aspects as highlighted in Section 2.

3.2.1 Experiential Learning Theory

ELT has its roots in the theories of John Dewey, Kurt Lewin, and others (Scarce, 1997). ELT considers experience as the central source of learning and development and consists of four stages as shown in Fig. 3.1:

1) Concrete experience (CE): perceiving new knowledge by experiencing the concrete through sensing and being immersed in real situations.
2) Reflective observation (RO): watching and reflecting on the learner's own experiences or that of others.
3) Abstract conceptualization (AC): analysing, thinking, or planning through a symbolic presentation.
4) Active experiment (AE): doing things.

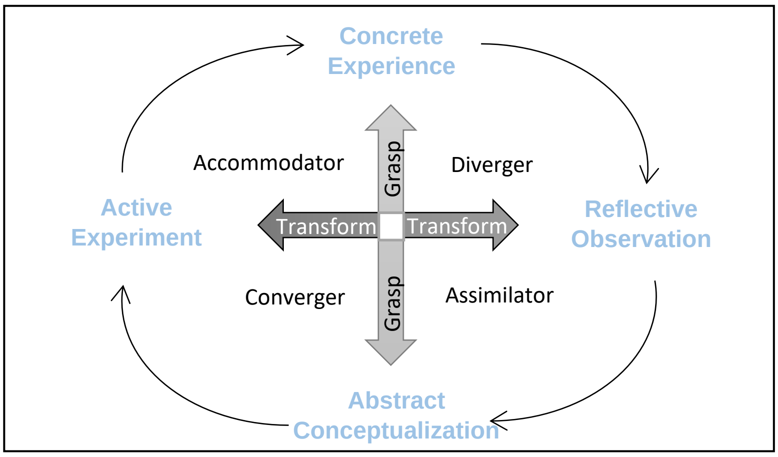


Figure 3.1: The ELT cycle & learning styles (Kolb 1984).

These four stages are performed in a cycle and can be started at any stage and repeated as needed. The student has to encounter all stages: experiencing, reflecting, thinking, and acting. The theory emphasizes the importance of experience in the learning process, such as laboratory sessions and fieldwork based on several foundational works (Lewin, Dewey's theory of experience, Piaget's cognitive development theory) (Kolb, 1993). Learning, defined based on ELT, is "the process whereby knowledge is created through the transformation of experience. Knowledge results from the combination of grasping and transforming experience"(Kolb, 1984). The four stages form two associated modes of grasping experience and transforming experience. The learner can start the cycle from any point but has to touch all four bases. ELT in practice gives the learner the choice in grasping and transforming their experience.

Grasping can be performed in two stages: grasping by experience (CE) or grasping by abstracting (AC). Transforming also has two methods: transforming by reflecting (RO) or transforming by acting (AE). The theory originally defined four learning styles that depend on the learner's preference to choose among the four stages: experiencing, reflecting, abstracting, or acting. This preference results from personality type, life experience, and cultural influences. The learning styles are: diverging, assimilating, converging, and accommodating.

3.2.2 Internal Economy

Games in general build on rules of play, and digital games consider game rules as mechanics that govern the relationships between gameplay components. Game mechanics are the detailed and hidden rules of games, and players do not need to know all the rules explicitly at the start of the game. Usually, there is a core mechanic in a single game which is the most important and impactful on aspects of the game. There are five types of game mechanics: physics, internal economy, progression, tactical manoeuvring, and social interaction (Adams, 2014; Adams & Dormans, 2012). The physics mechanic is about applying the science of force, gravity, and motion; however, the physics can be changed in the game environment by exaggerating sometimes. The internal economy mechanic handles the flow and transaction of game elements that are considered resources in gameplay, such as coins and lives. The progression mechanic controls the player's progression through the game's challenges and levels, where the player moves by unlocking enclosed areas of the game environment. The tactical manoeuvring mechanic focuses on the placement of units and the advantage that might be gained in each possible location. The social interaction mechanic governs the interaction between players. Digital games in general combine more than one game mechanic. However, the internal economy is the basic mechanic and most involved in designing digital games.

As mentioned previously, almost every digital game has an internal economy that is similar to a real-life economy. A general definition of an economy is a system that produces, consumes, and trades resources in measurable amounts. The internal economy of games manipulates many kinds of resources which can differ from what people are used to in real life. For example, resources in games include health, lives, and stars in addition to food and money. Rules of gameplay form part of the internal economy by governing the transaction and flow of resources in a quantifiable amount. Internal economies of games range from small and simple to more complicated and complex systems, the development of which is an important task in game design. The internal economy is important because it is a mechanism for defining and designing the game's rules in relation to resources. Three components form the internal economy: resources, internal mechanics, and the feedback loop.

Any object or material in a game that can be quantified numerically is a resource, such as money, enemies, ammunition, and energy. In games, players can control things by gathering, destroying, or producing different objects which formulate resources. Resources can be tangible or intangible. Any resource that has a specific location with physical characteristics in the game environment is tangible, such as any collected item, while intangible resources have no physical space to occupy, such as points obtained from collecting items. A particular number of resources need to be saved and stored in a container called an entity. For example, when a player collects gold, it will be stored in an entity (gold box), or an entity (timer) can be used to store time. An entity can be simple, storing one type of resource, or compound, where more than one simple entity is associated together. For example, in a racing game, a car includes wheels, fuel, and speed; each element represents a simple entity and together they form a compound entity.

In an internal economy, resources flow from one entity to another regarding four internal mechanics: source, drain, converter, and trader. The source mechanic produces new resources and stores them in entities, such as creating new dots in the Pac-Man game. The production of the source could be based on a condition, triggered by an event or automatically based on a time interval. Also, sources have a production rate, which can be fixed or variable depending on the time or amount of another resource. The condition, automation, and changing rate are concepts that are applied to all internal mechanics. The reverse mechanic of the source is the drain, where resources are consumed and removed from the game world permanently. The quantity of resources is reduced by a specific amount. On the other hand, the converter changes one type of resource to another, such as converting flour to bread. The converter mechanic also works at a specific rate; for example, one bag of flour could be converted into five loaves of bread. The trader mechanic exchanges two different resources between two different entities according to a specific rate. For example, a player can trade a shield to get a more powerful gun: no resources would be destroyed or produced.

In an internal economy, when a resource that results from a specific mechanic feeds back and affects the same mechanic at a later time in the game, this is called a feedback loop. For example, taking one piece of the opponent in a chess game will make it easier to take the next piece. A positive feedback loop applies when the effect of the loop becomes stronger in each loop. However, the positive feedback loop can cause a deadlock when the production of two resources is mutually dependent on each other. For example, in a construction mechanism, when building a stonecutter's hut in Settlers III, the stonecutter's hut produces stone and at the same time the player needs the stone to build the stonecutter's hut. The game starts with some stones but if a player uses them for other tasks before building the stonecutter's hut, then they could end up without enough stones to build the hut, which will result in deadlock. A positive feedback loop helps the player to win quickly when an important difference is achieved in skill or effort. On the other hand, a negative feedback loop stabilizes the production mechanism in the internal economy. For example, in a car racing game, the positions of players' cars appear to be attached to each other by a rubber band. No car will get too far ahead of the others or too far behind the rest. This can be balanced by powering up the slowest car with random power or increasing the difficulty of the leader car with blocks. It will increase the excitement by creating chances for other players to go into the lead.

3.2.3 Game Machinations

The game machinations framework (Adams & Dormans, 2012) is a tool to visualize game mechanics represented in resources along with their flow, in addition to transactions of the internal economy. This research utilises the game machinations framework to introduce learning aspects to game mechanics with the aim of linking them. It presents the link between game elements and educational elements, and these elements can be manipulated to reach the desired setting for designing GBL.

The symbols of game machinations are a way to facilitate and support modelling of the internal economy in a graphical representation. For example, entities that store resources are represented by an open circle (pool), while resources are symbolized by small coloured circles (coins) or as numbers. Another example is the source mechanic, which is represented by a triangle pointing upwards, and a solid arrow, which represents the flow of resources from a source to a pool entity. Table 3.1 provides a description of game machinations symbols and some machinations concepts used in this research.

Table 3.1: Symbols of game machinations.

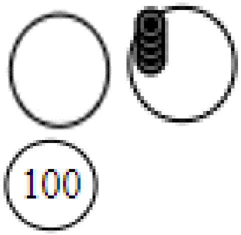
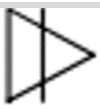
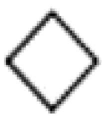
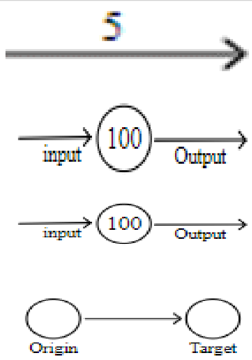
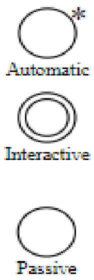
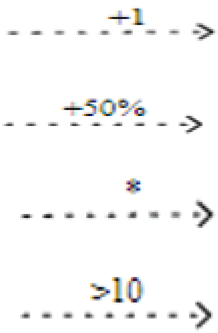
Symbol	*Description*
	**Pool** is an entity used to store resources. It is represented by an open circle, while resources are represented by a coloured small circle. Resources can flow from one pool to another at different rates. A large number of resources displays as numbers inside the pool.
	**Source:** creates resources.
	**Drain:** consumes resources.
	**Converter:** converts one resource to another.
	**Gate:** redistributes resources immediately without storing them. It is represented by a diamond and with one or multiple outputs. Its label shows a condition or probability of flow.
	**Resources connection** is a way to move resources in the machination diagram. It is represented by a solid arrow along with a label showing the rate of transferring resources in a one-time step. The arrow connects two symbols or nodes to move resources from one entity to another.

Table 3.1: Symbols of game machinations (Continued).

Symbol	*Description*
	**Activation mode**: the fire mode of a node in the machination diagram. The firing involves sending resources via source connections. **Automatic:** automatic firing in every time step. **Interactive:** firing in response to player action. **Passive:** firing in response to a trigger. These activation modes are applied to pools and other nodes such as a drain or source.
	**State connection:** represented by a dotted arrow to show the effect of the current state of a node on something else. There are four types of state connections. **Label modifier:** affects the value of the label associated with the target resource connection by the changes in the origin node. It connects the node to the label of the resource connection. **Node modifier**: the state change of the origin node alters the number of resources in the target node. It connects two nodes. **Trigger:** distinguished by an asterisk (*) in the label and activates the node or resource connection connected to when the origin node satisfies its inputs. **Activator:** connects two nodes, and its function is activation or inhibition of the target node based on the state of origin and condition in its label.
Flow rate	Resource connection moves resources at different rates, which is the number of resources that will be moved from node to node. The flow rate can be a number, percentage, or random (die symbol).
State change	The current distribution of resources between nodes in the machinations diagram (number of resources in pools)
Time step	Machination diagrams are time based and resources are moved and updated in each time step.

3.3 Research Design

This research includes two studies to answer the research questions and evaluate the proposed conceptual framework: preliminary study and main experimental study. After constructing the conceptual framework, a first step was established by conducting a preliminary study to explore the experts' perceptions (game designers/educators) of usability and usefulness. The preliminary study collected data via an online questionnaire - Bristol Online Survey. The second step involved designing and developing a prototype based on the proposed framework to create VFTG. Then the main study which is a quasi-experiment is conducted to measure variables such as learning performance to evaluate the effectiveness of the proposed framework. Different statistical analysis tests are performed on the collected data from both studies (preliminary & main) with the aim to answer the studies' questions accurately.

3.4 Research Elements

The research studies are formed based on several elements, and these elements are introduced briefly in this chapter and explained in detail later in the related chapters (5 and 7). Both studies consist of participants, instruments, variables, and data analysis tests. Each element is important to be selected carefully to answer the research questions and validate this proposed work.

3.4.1 Participants

The preliminary study of the conceptual framework is designed to explore the experts' perceptions of usability, usefulness and connection. The goal was to obtain experts' feedback in order to identify issues before starting the implementation of the VFTG prototype. The target participants are game designers/educators from universities and schools. Twenty-three voluntary respondents answered the requests which sent via email. The main study involved the VFTG prototype that is designed based on learning content from the recent National Curriculum in England for geography Key Stage 3. It is a quasi-experiment that operated in a local school in Durham city. The conceptual framework is developed to support the process of designing VFTG in general regardless of the type or age of participants. However, the participants of the main study were chosen to be the audience who would most benefit from the VFTG prototype as the learning content is aimed to benefit secondary school students. A total of 60 learners participated in one of two groups, 30 (50%) in the experimental group learnt by playing the VFTG and the rest in the control group learnt by conventional method (traditionally taught by a teacher). The year group is 7 for both groups and the age is 11-12 years as normally expected for this year group. The learners of both groups have the same level of ability of comprehension.

3.4.2 Instruments

For the preliminary study, one instrument was applied to collect data which is a questionnaire. Two types of data were gathered by the questionnaire, qualitative and quantitative. The questionnaire aimed to answer questions about the connection, usefulness, and usability of the framework. The connection was evaluated by ten questions, while the usefulness was evaluated by five questions and all the questions required answers based on a five-point Likert scale (strongly Agree=5, Agree=4, Neutral=3, Disagree=2, strongly Disagree=1). The usability measured by The System Usability Scale (SUS) (Brooke, 1996), which is a well-established questionnaire for measuring usability. SUS contains ten items of attitude Likert scale to subjectively evaluate the usability in a global view. The experts' suggestions are captured via two open-ended questions. In addition to personal questions that focus on knowledge of game designing/FBL, working experience, and type of working organization. The questionnaire and summary of the proposed framework were provided via Bristol Online Survey and invitation was sent via email and followed by remainders after one week.

While the main study facilitated several instruments to collect data as follows:

Pre-test and post-test employed to assess learners' learning performance in volcanos - natural hazard topic. The pre-test was delivered to learners before the experiment and the post-test was performed after four weeks of completing the learning process. Both tests include 10 multiple-choice items that followed the revised Bloom's Taxonomy. The post-test consists of the same 10 test items in a different order from the pre-test.

Intrinsic Motivation Inventory (IMI) is an adjustable measurement tool that identifies persons' levels of intrinsic motivation associated with a particular activity such as playing a game. The tool allows the discovery of the VFTG's effects and provides indications on intrinsic motivation at a deep meaningful level by evaluating variables that are assumed to accelerate the process of intrinsically motivating feeling. IMI (McAuley, Duncan, & Tammen, 1989; Plant & Ryan, 1985) assesses motivation regarding to 7 Subscale (interest/enjoyment, perceive competence, value/usefulness, felt pressure and tension, value/usefulness, perceived choice, and relatedness). Only the applicable subscales to this research are selected. The modified version of IMI utilised in this research contains 19 statements scored on a seven-point Likert scale from 1) not at all true to 7) very true. The preferred subscales are: Interest/Enjoyment - 7 statement, Pressure/Tension - 5 statement, and Value/Usefulness - 7 statement.

igroup presence questionnaire (IPQ) (IPQ, 1999; Schubert, Friedmann, & Regenbrecht, 2001) is employed to assess the experience of player presence in the VFTG. The questionnaire contains a general statement and 3 factors, which can be considered as fairly separated factors. These three factors evaluate the following:

1. Spatial Presence (SP): the sense of being really present in the virtual world.
2. Involvement (INV): the interest in the virtual world and the involvement in the experience.
3. Experienced Realism (REAL): the player's experience of realism in the virtual world.

The general statement assesses the common sense of being there, and has a huge effect on all three factors, especially on SP. Statements measured on a seven-point Likert from 0 to 6 with several meanings such as fully disagree/fully agree or did not feel/felt present.

Gameplay data

The gameplay data was collected for the experiment group based on specific indicators that developed as part of the evaluation model (see Chapter 4, Section 4.2.2.3) of this proposed framework. The log data was stored on a server (www.interserver.net) as a simple database. The evidence of constructing knowledge and improving skills can be measured by the collected indicators that capture the player's interactions in the VFTG. The data (indicators) is collected for learning tasks that are included in the VFTG prototype and explained thoroughly in Chapter 6 and Section 6.4.

3.4.3 Study Variables

Several variables are selected with the aim to understand and evaluate the effect of the proposed conceptual framework. These variables are learning performance, learning experience (motivation and presence), and gameplay data. The main goal of this research is to enhance learning and for that the learning performance variable is expected to be measured. Learning performance is an important impact of GBL. Therefore, to evaluate the proposed conceptual framework by utilising the VFTG prototype that was designed based on, learning performance is selected to be measured. Motivation is one of many variables that studies found to affect students' performance (Malone, 1981; Ryan & Powelson, 1991) and at the same time is important of keeping players playing digital games (Ryan, Rigby, & Przybylski, 2006) and it has been utilised to evaluate GBL and SGs (Eseryel, Law, Ifenthaler, Ge, & Miller, 2014; Kim, Ke, & Paek, 2017; Martín-Hernández et al., 2021). Motivation is an essential variable to produce effective learning and playing and for that it is selected to be measured. Presence is one of the elements that concern the designing and evaluation of virtual environments such as VFT (Grassini & Laumann, 2020; Skarbez, Brooks, & Whitton, 2017). As mentioned in the literature review chapter (Chapter 2, Section 2.3), that one of the drawbacks of RAFTs is missing the feeling of "being there" (Stephens et al., 2016) when students watch a live podcast of their teacher out in the field site alone or with a small group of students. Also, the connection between presence and human performance such as learning performance is considered to be positive (Grassini & Laumann, 2020; Stevens & Kincaid, 2015). So, the selection of the presence variable is based on its importance to the designing of virtual environments in general and its positive effect on learning performance. The gameplay data are selected regarding the interactions and indicators that are defined/identified in the evaluation model (see Chapter 4, Section 4.2.2.3), and designed/collected based on the VFTG prototype (see Chapter 6.3.2.4, Section 6.4).

For evaluating the learning performance two main variables are analysed: the time effect (τ) and the intervention effect (η). The control group determines the adjusted odds ratio of responding correctly with regard to the effect of time only (e^{τ}). On the other hand, the experimental group deals with both the effect of time and intervention ($e^{\tau+\eta}$). Therefore, the independent effect of the intervention can be defined as

$$\frac{e^{\tau+\eta}}{e^{\tau}} = e^{\tau+\eta-\tau} = e^{\eta}.$$

Regarding IMI, the main variable is the interest/enjoyment subscale score in addition to the other subscales' scores. While regarding IPQ, all the factors' scores are considered equally important. In addition to these main variables, secondary variables are analysed to find their associations and effect on the main ones. The secondary variables are: gender, post-test, difference score (post - pre), and frequency of playing video games. The frequency of playing games consists of five different levels as follows: every day, 3-5 times per week, 1-2 times per week, not very often, and not at all.

Regarding the gameplay data, the variables reflect the indicators that are defined based on the proposed conceptual framework and identified during the designing and implementation process. These variables:

- Capture the players' interactions that discover gained or employed subject knowledge as a result of performing game tasks (knowledge indicators).
- Distinguish the players' interactions that show acquired field-based knowledge (Field-based indicators). This type of variable captures the player's ability to observe and also saves information about the player's path movement in the VFTG environment.
- Recognise the players' interactions that identify developed skills (skill indicator). This type of variable measures the level of proficiency of decision-making skills.

3.4.4 Data Analysis Tests

The statistical analysis test that is employed to analyse learning performance based on pre-test and post-test is the linear logistic test model (LLTM). LLTM is a generalization of the Rasch model (Fischer & Molenaar, 1995). LLTM can be utilised to evaluate the change in learning performance and to distinguish the item difficulty that relative to cognitive operations, or any other features of the item material. The item parameter of a test item at the second time ($T2$) in quasi-experiment will be determined as the sum of the item difficulty parameters at the first time ($T1$) plus the parameters of the change that associated with the intervention. By applying LLTM, the effects of time and intervention can be recognized. In addition, LLTM is appropriate for the pre/post-test answers data type which is dichotomous (correct or incorrect). The analysis of IMI subscales and IPQ factors is achieved by the explained process of the original authors which included the procedures of determining the score of each regular statement and reverse statement. Then calculating the scores of each subscale individually by computing the mean of statements' scores on that subscale.

The counts and percentages are employed as a part of the descriptive analysis to summarise information about some study variables such as the percentage of failed/passed learners. The median is calculated to measure the central tendency, which is the middle point for a set of data after arranging them in order of a scale. The median is preferred where it is less affected by outliers and skewed data. In addition, the interquartile range (IQR) is utilised to measure the variability based on dividing the dataset into quartiles. Specifically, IQR is the amount of spread in the middle (50%) of the dataset, where the ordered dataset is divided into four equal parts. Each part is called a quartile and the first quartile is the middle point in the first half of the ordered dataset, the third quartile is the middle point in the second half of the ordered dataset and the interquartile range is the distance between the first and third quartile. The IQR shows how spread the data is and unlike the standard deviation it is also less affected by outliers or skewed data. Cronbach's alpha (α) is applied to measure the internal consistency (reliability) which is commonly used with Likert-scale questions in a questionnaire and provides an overall reliability coefficient. It assesses the consistency of statements or questions verifying that all statements positively and strongly correlate with each other to be a consistent measure of a concept. The Cronbach's alpha is obtained for both pre-test, post-test of each group, IMI subscales, and IPQ factors.

Spearman's rho (ρ) correlation coefficient is utilised to determine the direction and strength of the monotonic relationship between two ranked variables where monotonicity is less restricted than the linear relationship. In fact, Spearman's correlation is the nonparametric adaptation of the Pearson correlation. Spearman's rho (ρ) correlation coefficient is calculated to define the relationships between several study variables such as between IPQ factors and post-test and the difference scores. The Mann-Whitney U test is applied to assess whether two samples probably come from the same population. It tests if there is a difference in a dependent variable regarding two independent groups. It evaluates if the dependent variable's distribution is identical for two groups. It is regularly thought of as the nonparametric alternative to the independent t-test. In this research, the Mann-Whitney U test for independent samples is applied to determine the effect of gender on IMI subscale, and IPQ factors.

The analysis of gameplay data can be considered as a way of measuring the learning performance of players during the VFTG. Also, a comparison between the learning performance during the VFTG and the learning performance that was measured after completing the VFTG via post-test could be performed as future work.

3.5 Research Ethics

All questionnaires and tests were collected anonymously without any private or identity information. In addition to obtaining the research ethical approval from the Ethics Representative of the Department of Computer Science. All data are assured to be exploited for this research only and will not be shared with the public. Also, the researcher gained the enhanced Disclosure and Barring Service (DBS) to work directly with children (secondary school learners) in a controlled setting to undertake a quasi-experiment. All participants are invited to participate in the experiment willingly with the attendance of their teacher.

3.6 Summary

- This chapter provides a part of the answer to the second question of this thesis. Some of the strategies that utilised in this thesis to connect FBL aspects to game aspects (ELT, internal economy mechanic, and game machinations).
- The research design includes two studies: a preliminary study where the experts' perceptions of the conceptual framework's connection, usefulness, and usability are explored, and a main study where a quasi-experiment is conducted to evaluate the effectiveness of the proposed framework.
- The data collection tools that employed are questionnaires, pre/post-tests, and game log data.
- The statistical analysis tests that were performed are counts, percentage, median, IQR, Cronbach's alpha, Spearman's rho, Mann-Whitney U, Kruskal-Wallis and LLTM.

Chapter 4

4 Conceptual Framework

4.1 Introduction

The process of creating digital games involves both design and development stages; however, this thesis focuses on the design stage of VFTGs. A general overview of the VFTG creation process is presented in Fig. 4.1. The process starts with designing the VFTG based on a conceptual framework that aims to connect FBL theories and pedagogies to game elements and mechanics, which is the core purpose of this thesis. The proposed conceptual framework is concerned specifically with the learning task, skill, and assessment. The design stage of VFTGs also manages the provision of feedback and tracking of the progress of players' learning. The progression of the player and acting on feedback shapes the adaptation of learning task difficulty, in addition to the learning style and prior knowledge of the player.

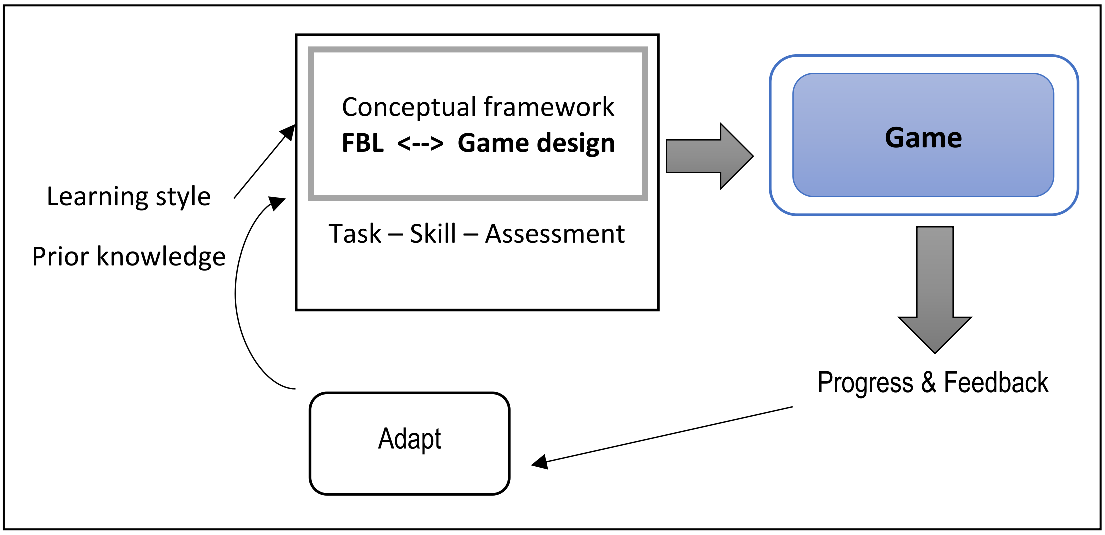


Figure 4.1: VFTG creation process cycle.

The conceptual framework of VFTG design is supported by learning and game design aspects (internal economy and game machinations). ELT, internal economy, and game machinations were introduced briefly in Chapter 3 to establish the essential aspects of this research. The following sections explain the components of the conceptual framework: section 4.2.1 describes ELT modelling, where ELT is quantified as the internal economy in the form of game machinations; section 4.2.2 defines game modelling which is achieved via four steps (matching scheme, task model, evaluation model, and skill modelling); and section 4.2.3 explains world modelling.

4.2 The Framework Components

The framework connects FBL to game design (game elements and mechanics), enabling the design of VFTGs in a process consisting of a number of steps. It includes three links from main FBL aspects to game design components and is a cyclical process of designing and improving as shown in Fig. 4.2. The three links are:

- ELT modelling maps the theory into the internal economy mechanic.
- Game modelling links the learning process to gameplay.
- World modelling connects the field environment to the game environment.

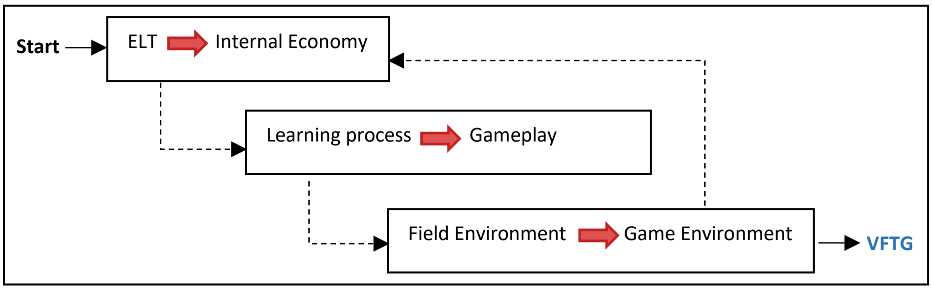


Figure 4.2: The three links.

4.2.1 ELT Modelling

Game machinations are applied in this research and simplified to be readable and understandable to educators as well as game designers. The internal economy is used to quantify ELT concepts to link them to game design. ELT is the progression of learning through experience and reflection on doing. For each stage of ELT, resources are defined along with a suitable internal mechanic and feedback loop when needed. These three components (resources, internal mechanic, feedback loop) transfer the theory into the game's internal economy, which is a core mechanic of designing games, as shown in Fig. 4.3. The next section provides the modelling of ELT as the internal economy. Each stage of the theory is modelled to form a building block of experiential learning.

It can be viewed as an offer by educators/game designers and a request from learners, while the purpose is to reach the stage of purchasing by learners, as shown in Fig. 4.4. The game designer/educator offers learning by selling (knowledge, skill, or experience), while the player demands fun and will play or "purchase" what has been offered when he/she finds enjoyment, engagement, receives valuable goods (grades and recognition), or resolves a conflict to test his/her knowledge of the information needed to overcome challenges. In an economy, goods are purchased for money; however, in GBL the game designer/educator expects players to purchase learning by spending time playing the game and gaining knowledge/skills.

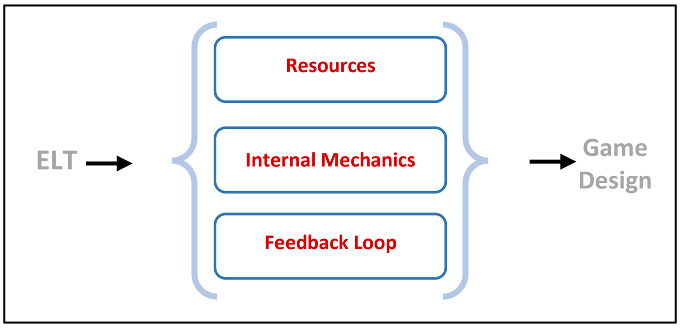


Figure 4.3: Modelling ELT as internal economy.

It is a model of the learner's progress in reaching different stages of the ELT cycle while performing learning tasks. By completing a stage, the player's knowledge/skill will be developed to show progression in performance. Four resources are defined, each of which represents a different level of achievement and is associated with a specific stage of the ELT cycle. Levels of achievement are ***level1***: observing, ***level2***: reflecting, ***level3***: synthesizing, and ***level4***: doing. Internal mechanics (source and converter) are used to show the flow and transaction of these resources, which are the different levels of achievement, from one stage to the next one in the cycle. Solving a learning task and constructing new knowledge in steps by going through the cycle can be repeated via a feedback loop to improve performance in the next cycle by acting on the learning feedback.

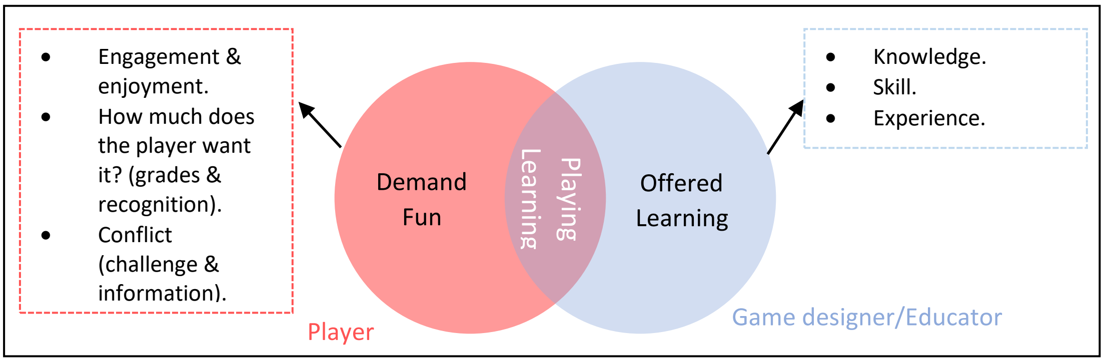


Figure 4.4: Purchasing, playing, and learning.

The first stage of ELT is the concrete experience (CE), which involves gaining knowledge through new experiences or recalling previous experiences. This is a simple internal mechanic where the player grasps knowledge, and it is abstracted as the interactive source *Task* to complete a task. The difficulty of the task is represented as a gate with the *player* skill symbol and the probability of successfully producing *level1* as *p1*. For example, *p1* could be 50% chance of completing the task successfully based on the player's ability level and producing an observation resource to be stored in the *level1 pool*, as displayed in Fig. 4.5. The resource of the first level of achievement is defined as observation because in the first stage of the cycle, the player is expected to have a new experience and develop knowledge via exploring and observing the virtual field.

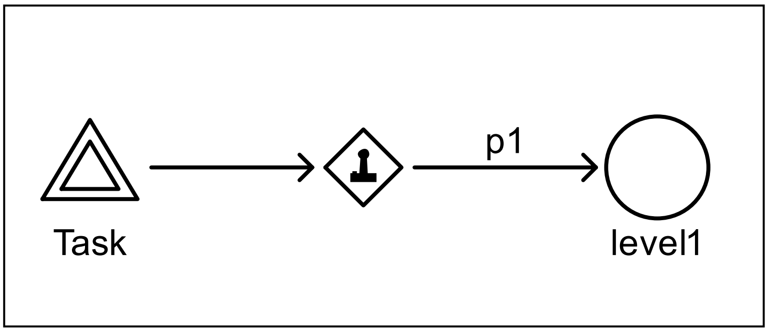


Figure 4.5: Concrete experience (CE).

Reflective observation (RO) involves the player thinking and reflecting on their experience from the first stage of ELT. It is a mental activity but can be encouraged by activities such as discussion. In Fig. 4.6, reflection represents a source, *Reflect*, which can be generated in one of two different situations: whether or not the player completed the task in the first stage successfully. In the first situation, as a resource is stored in the *level1* pool, a trigger state connection from the *level1* pool to the source *Reflect* will trigger reflection, as the player is expected to reflect when they successfully complete the task. In the second situation, which applies if the player could not complete the task, the gate guarantees the redistribution of the player's experience to reflect by *p2*. Where *p2* represents the probability of player's ability to reflect when the task is not completed successfully, showing that he/she still learned something, even from mistakes. By the end of the second stage, the player should produce resources of reflection, which are stored in the *level2* pool.

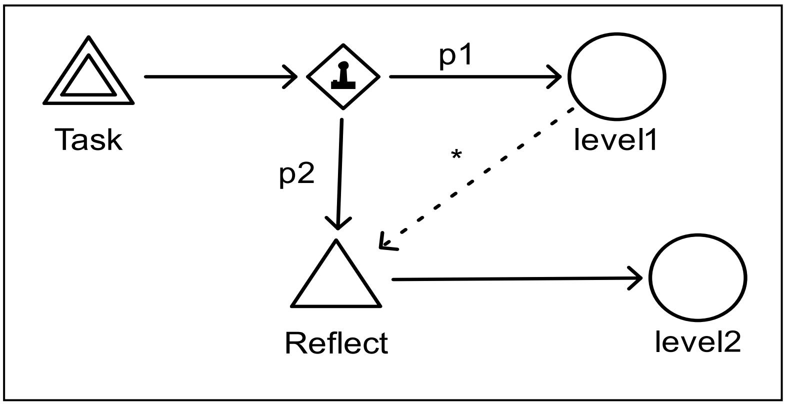


Figure 4.6: Reflective observation (RO).

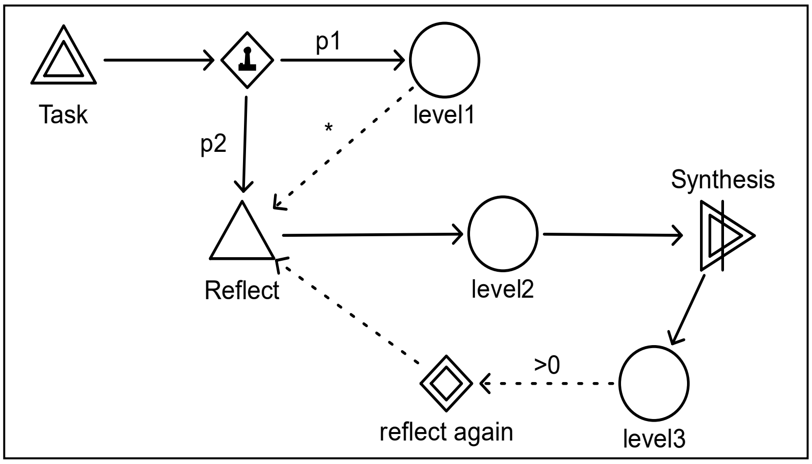


Figure 4.7: Abstract conceptualization (AC).

The third stage, abstract conceptualization (AC), involves synthesizing ideas/conclusions, hypothesizing, and planning to experiment in the next stage. This process can be done via a loop of synthesizing and reflecting again until the final plan is formed, as shown in Fig. 4.7. The loop starts from the interactive converter *Synthesis*, which converts *resources in level2* to synthesize and plan and then stores them in the *level3* pool. Storing resources in the *level3* pool will activate the interactive gate (*reflect again*) by the activator state connection and the condition specified on the label as (0>). At this point, the player has the choice to click on the (*reflect again*) gate or move on to the final stage. If the player chooses to click on the gate, the source *Reflect* will be triggered to generate more resources to be stored in *Level2* pool and then can be synthesized again.

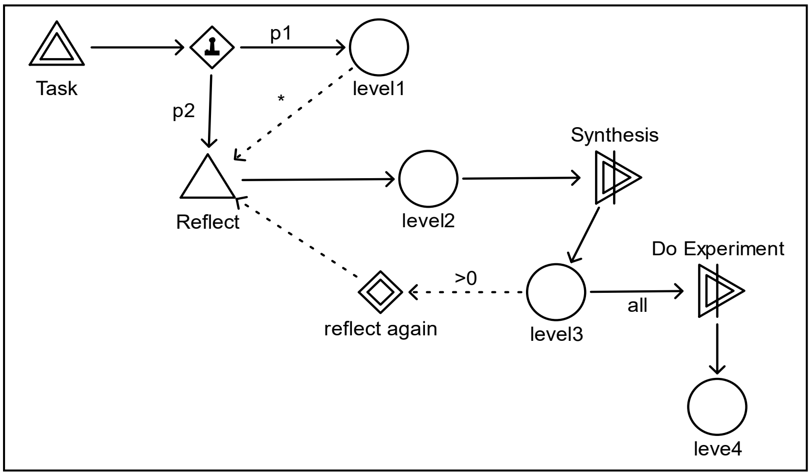


Figure 4.8: Active experiment (AE).

In the final stage, active experiment (AE), the player needs to practise what he/she learned from the previous stages. In Fig. 4.8, this final step is abstracted with the interactive *Do Experiment* converter, which represents the action of the player undergoing a new experience based on the synthesized ideas/conclusions and plan from the previous stage – *Level3* resources. This will result in producing new knowledge stored in the *Level4* pool after a full ELT learning cycle.

At the end of the cycle, the player is expecting feedback as a result of the evaluation to start a new cycle. The source *Generator* in Fig. 4.9 symbolizes feedback provision, which will be triggered automatically by storing resources in *Level4*. The generated feedback will be stored in the *Feedback* pool. The player has to act on the provided feedback by clicking on the interactive converter *Act.* The converter *Act* will produce *Action* resources, which means that the player utilised the feedback to improve or fix something in the task performance.

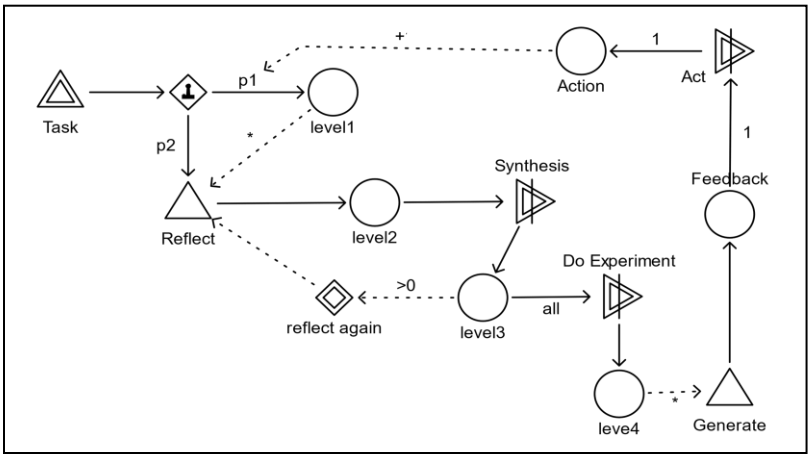


Figure 4.9: Feedback loop to the next learning cycle.

Working through the whole cycle and acting on the feedback will improve the player's knowledge and skill, which will increase the probability of completing the task successfully in the next cycle. This is achieved by a label modifier with (+), where the percentage will be increased by some value as decided by game designer/educator automatically each time resources are stored in the *Action* pool.

Each stage of ELT cycle forms a building block that can be facilitated in the process of VFTGs. The feedback provision and evaluation will be extended in the following section.

4.2.2 Game Modelling

The game modelling helps game designers/educators to connect learning process to gameplay through four steps. The first step involves utilising the matching scheme tool to shape the VFTGs structure by adapting the FBL structure to the virtual environment context and then linking it to suitable game elements. The second step shifts the designing process to focus on learning tasks via connecting learning elements (goal, objectives, outcomes, assessment methods) to game mechanics and selected game elements from the matching scheme. The third step develops the assessment via the evaluation model to benefit from the characteristics of VLEs. The fourth step focuses on a specific type of learning task which involves high-order skills such as decision-making skill and teamwork skills. The following subsections will explain each step thoroughly.

4.2.2.1 Matching Scheme

The matching scheme in Table 4.1 is a tool created as a part of the conceptual framework to support game designers/educators to transform the FBL structure into the virtual field structure in the game environment. It is a tool used to plan and define the VFTGs structure and link it to the required game elements. The tool includes three phases formed based on the FBL structure with adaptation to fit the VFTG environment: pre-, during, and post. Each phase is divided into a number of steps and each step is linked to several game elements to help develop the field trip (FT) experience in VFTGs. It is recommended to not ignore any of the steps, but the designer can be selective in which game elements are included within each step to match the learning objectives, and the steps of each phase are not required to be followed in order. The matching scheme is illustrated in table form to be efficient and yet easy to use by both educators/game designers. Applying the matching scheme would further develop and link FBL to the game design process. In order to utilise the matching scheme effectively, the following sub-section describes the game elements that can enhance the design of VFTGs and defines the relationships between them.

Game Elements of VFTG Design

The core game elements that can be applied to design VFTGs based on the conceptual framework will be discussed and explained in relation to FBL. Fig. 4.10 presents the relations between interaction and other game elements included in the framework. The game elements: control of choices, narrative, challenges, identity, and presentation contribute to creating opportunities for interactions in the VFTG environments which mimic the learning scenario in a physical FT. Interaction also results in entering the flow state which, along with other elements, produces a feeling of immersion/presence and motivation.

Central to this framework is ***interaction***, and the reason for this is the effectiveness of FBL grounded in interactions with the field environment. In general, learning is prompted by the interactions of the learner with the learning context, which includes the environment, peers, and learning materials. Designing an interactive VFTG environment invokes the development of high-order skills (analysing, problem solving, creativity, collaboration, and teamworking) while learners are immersed in the gameplay. The interactive learning environment includes different types of interactions that can be categorized into *social, environmental, and learning resource* interactions. Each type of interaction should be employed while designing the VFTG as required by the learning task. Social interactions promote communication between the player and peers when a VFTG is designed in a multiplayer format or designed to promote communication with NPCs. The VFTG environment along with players can be inhabited by NPCs who play different roles such as a guiding teacher, or supervisor who will be helpful, especially in a cook's tour and teacher-led VFTs. Interaction with the environment means the ability to move and edit game objects. Interaction with learning resources (written materials or video clips) encourages exploration and discovery, which leads to self-regulated learning. VFTGs should be designed to create opportunities for interacting within the virtual field environment.

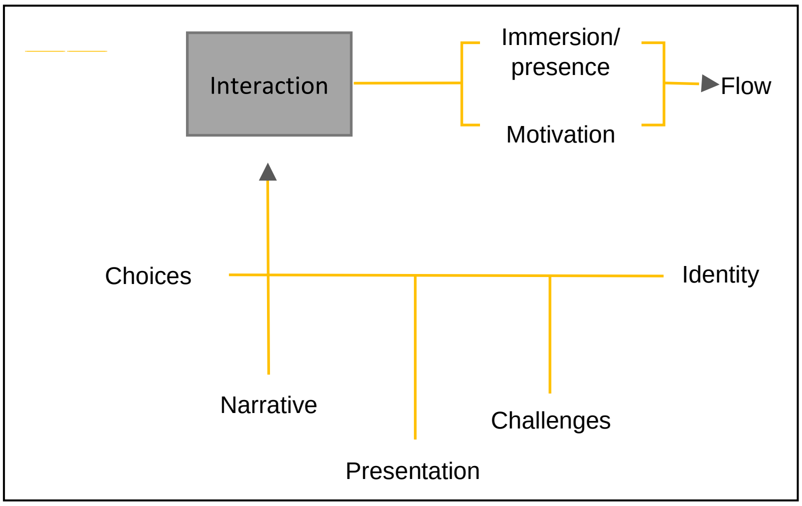


Figure 4.10: Game elements of VFTGs.

Interaction leads to ***control choices*** for the learner/player to manipulate objects or take a different path to complete a task. Control choices start with selecting an avatar and deciding on which role-playing character to impersonate. This is considered a way of allowing players to distinguish their own ***identity*** in the VFTG world. The feeling of control enhances the chance of players reaching the state of flow through engagement. Interactions with learning materials through observing, discovering, analysing, and making decisions encourage players to take an active role in their learning and support connecting learning content to practice in virtual field environments.

In addition, the ***narrative*** offers the player opportunities for interactions by providing choices that could alter the nature of the narrative, leading to different paths for doing learning tasks and more engagement with the gameplay. Attaching stories to VFTGs is a fundamental element of game design that motivates the learner by integrating ***challenges*** and learning tasks within the gameplay. A strong structured narrative gets the learner involved and immersed/presented within the gameplay. The narrative can be as simple as a background story of game missions such as ghosts chasing the player in Pac-Man. A story helps learners to structure and recall their own experiences and can represent reality or fantasy. In addition, learning content can be included within the story events, which will help players to learn in an immersion/presence way. Integrating narrative within the gameplay requires the application of narrative devices such as a backstory and cut scenes. The backstory forms the background of the game story, and cut scenes are elements that reveal pieces of the story during the game. The back story and cut scenes must be connected to the FT experience; for example, the story for a virtual trip to a historical site should match the historical timeline of that site to encourage players to live the experience. Narrative devices can take different forms, such as text, audio, image, or communication with an NPC. Integrating the narrative establishes possibilities for inquiry, taking an active role, and analysing.

Presentation in the game environment generally takes the form of multimodal media (motion, image, text, audio, and video). This multimodal presentation is essential to VFTGs because it helps to create an environment that represents the physical field. In addition, the multimodal presentation can be used for more than creating an amazing VFTG environment; it also provides a rich source of information and data. This is the case with a physical FT where learners interact with the environment (museum, zoo, theatre, or a field at the back of a school) to gather information (data), analyse it, form a hypothesis, and test it. Therefore, the environment is a main source of knowledge in VFTGs.

Based on the ***flow*** experience concept (Csikszentmihalyi & Csikszentmihalyi, 1975; Kiili, 2005), the balance between task challenge and the player’s skill level increases the opportunity for experiencing flow. The task should not be considerably less challenging than the player’s level of skill, which could lead to boredom. On the other hand, a task with a greater challenge than the player’s skill level would cause anxiety. The balance application is discussed in the evaluation model. Motivation and immersion/presence are expected with flow achievement. The explained game elements are essential for designing VTFGs, and game designers/educators can utilise them in forming the FT experience by applying the matching scheme.

Matching Scheme Tool

The matching scheme (see Table 4.1) consists of phases (pre-, during, and post-) which are linked to game elements. The post-phase usually takes place days or weeks after completing the physical FT, which could affect the assessment and reduce the benefits of feedback. This proposed conceptual framework sees assessment as an integrated part of VFTGs to ensure the effectiveness of feedback and adaptation.

The **pre-phase** consists of four steps: start, pre-task, complexity of task, and skill. In the start step, *learning style* and *prior knowledge* can be identified to help to adapt the VFTG to each individual player; this is applied to the start point of the VFTG to support differing the playing experiences based on that. This information can be added to the player profile at the beginning of the VFTG. Preferably, the learning style of the player should follow the learning style inventory (LSI) associated with ELT. Another possibility if the learning style and prior knowledge are not available is to set the start point to a default setting and adapt after the first learning task.

Also, in this phase, the player should be allowed to select his/her *identity* to match the role he/she is playing in the VFTG. Permitting the player to select their identity provides *control of choice* and connection to the *narrative*, which leads to motivation and immersion/presence. In a physical FT experience, learners can relate their identity to some degree to the task they perform, such as investigating/investigator, collecting data/collector, and analysing/analyser, but VFTGs create stronger identity fulfilment. The pre-task step is about preparing the learner to do the task through planning, practice on a similar task, or giving a model showing the solution. This can be achieved by providing a *tutorial* level or *instructions* after selecting an identity. In addition, introducing a compelling *narrative* or theme from the start of the VFTG would lead to a feeling of immersion/presence and motivation. All of these elements (tutorial level, instructions, or narrative) should be supported by *multimodal presentation* such as sound, text, and motion that is suitable for the learning content.

Setting task complexity is an important preparatory step for the following phases. In VFTGs this can be manipulated by two variables: location and time. The location variable defines the place that players will visit in the VFTG as one location (*one level*) or more than one location (*levels*). For example, a VFTG of a museum can be limited to a single exhibit room or extend to multiple areas inside the museum. This can be mapped to one scene or level for each location, with the condition of success or completion required before going to the next location (scene/level). The *time* element can be added to form a further level of complexity with a variant amount of time for a scene/level, which could mean more tasks. Further details about the learning task variables are provided in the task model.

Table 4.1: Matching scheme.

Pre-	**During**	**Post-**
Start	Concrete experience - CE	Assessment
Adapting : Learning style Prior knowledge Player profile Character: Identity Choice Narrative	Explore: Narrative Mystery Interaction Chance Multimodal Presentation Emotion	Evaluate: Points Provide feedback: Multimodal Presentation Player's acting: Choice Action Recognition: Badges Calling Cards Skin
Pre-task	Reflective observation - RO	
Tutorial Instruction Narrative Multimodal Presentation	Reflect: Challenge Interaction Choice	
Complexity	Abstract conceptualization - AC	
Single location: One-level Multi-location: Time Levels	Synthesizing: Interaction Resources Challenge	
Skill	Active experiment - AE	Progression
Teamwork: Multiplayer Competition Collaboration Decision making: Time Choice Resources	Doing: Challenge Interaction Choice Uncertainty Consequence Time	Progression Bar Unlock Content

The last step in the pre-phase involves identifying skills such as teamwork or decision-making. The skill should be selected in relation to the learning content and can differ from one VFTG to another. The skills can be introduced to the player via the pre-task step. Each skill requires different game elements, and two examples are provided where the designing of teamwork skill tasks must include *multiplayer*, *competition*, and *collaboration* elements. The decision-making skill tasks require a different set of game elements such as *time* and *resources* to weigh the decision and *control of choice* to select the decision.

In the **during phase**, the learning task, which is essential to learning and gameplay, should be defined based on the required game elements. The learning task is divided into subtasks where each subtask forms one of four stages (CE, RO, AC, and AE). These subtasks support a progression mechanism in guiding the player to go through the entire four stages one by one regardless of the start point (ELT requires stages to be followed in sequence and the learner can start from any stage). Feedback can be provided after each stage or at the end of the ELT cycle. After utilising ELT modelling to define each stage, the educator/game designer can facilitate the matching scheme to link the appropriate game elements to each individual stage.

Interaction is necessary for all ELT stages, especially CE and AE. Interaction supports exploring and discovery, which are most needed in the CE stage. Involving some elements of *mystery* in CE stage will trigger curiosity and motivate the player in exploring and discovering the VFT environment. Mystery presents the need to bridge the gap between the known learning content from previous learning experiences and the unknown of the new experience. *Multimodal presentation* is important in designing learning subtasks at the CE stage; it motivates players to explore while interacting with the learning environment to observe the surroundings and find information and collect data. This is in addition to presenting learning materials in multimodal ways. *Narrative* in the CE stage of the theory creates a motivated and immersive context for the subtask and acts as a hook for the player to go through the following stages. Also, *narrative* supports the availability of more *choices* being given to players, which could lead to changing the narrative and result in more choices; it is a positive feedback loop. Attaching *emotions* to the learning experience enhances players' recall of knowledge and skills. Integrating emotions such as happiness, the rush of adrenaline to survive, or even sadness into the learning tasks of VFTGs would strengthen the connection to the learning experience.

The RO stage requires *control of choice*, where the reflection on the previous subtask varies from one player to another. Also, the *challenge* to motivate the player to reflect is necessary to guarantee progression to the next stage. The *choice* represents the different reflection for each player; this is an essential component so the player should be challenged to do it. It is a balance between freedom of choice and the necessity of reflection. As mentioned above, *interaction* is required in all of the ELT stages. Interaction in RO helps the player to find more information to support reflection or to re-interact with CE events to ensure the accuracy of reflection. Interaction can be one of three types (social, environment, and learning content) and all three types can be applied in one subtask or just one of them.

The AC stage is similar to the RO stage, consisting of mental activities which need to be challenging to encourage players to perform them. *Interaction* supports the process of synthesizing concepts, models, or ideas; it can be achieved by *challenging* the player to abstract a definition of a concept he/she experienced in the first stage via communication with a peer in a multiplayer VFTG, to chat with an NPC in a single-player VFTG, or formulating a model in an electronic notebook/table. The *challenge* can be in a form of a puzzle and the answer is the expected synthesized concept/theory. The player may seek more *resources* to perform the synthesizing process. These resources can be in the form of observed information from the VFTG environment, such as observing the colour of plants, the behaviour of animals, or the production process in a VFTG of a factory. The player's attention can be directed to this information via virtual feedback. Also, the resources can be of any type, such as video clip learning materials or text embedded within the gameplay.

The AE stage depends on providing a *challenge* to apply what has been experienced and learned from the previous stages. Therefore, the AE stage requires giving the player some *control of choice* to perform the subtask based on his/her understanding of the three previous stages. AE focuses on doing and acting, which definitely requires *interaction* to test the synthesized concepts/theory and should be designed to evoke different types of interaction as necessary. The *uncertainty* of what comes next in the AE stage drives the learning experience with the element of *chances* to analyse and question each step of solving the subtask before doing it. It is reasonable to consider the *time* element more carefully in the AE stage than in the previous three stages, where it was better not to rush the player through exploring, reflecting, and synthesizing. In addition, an element of *consequences* (positive or negative) will be helpful in this stage to guide the player through the experience instead of waiting for the feedback of the completed cycle.

The **post-phase** focuses on assessment and its related concepts (feedback and progression). Formative assessment plays an important role in the learning process because the player is expected to employ the feedback to improve learning in continuous ELT cycles by bridging the gap between actual performance and the desired one. Assessment and feedback can be provided for each subtask or after completing one ELT cycle. *Experience point* (XP) is a good example of the game element that can be utilised as a unit of measurement to quantify the player's progression and would reflect the grades of evaluation. Recognition can be achieved with game resources or currencies that cannot be exchanged such as a *trophy*, *badge*, *skin*, or *calling cards*; the purpose is to utilise game elements to acknowledge the player's progression. Recognition among players encourages social interaction and competition. Evaluation should be combined with feedback and a wide range of game elements can be utilised to provide feedback with *multimodal presentation* (audio, text, image, video), such as an NPC giving an audio message as a hint. In VFTGs, feedback can be embedded into the VFTG environment to capture the player's attention and keep him/her engaged. This can take the form of changing the colour of an object (animal, tree, or artefact in a museum) or animating the wind in a specific direction as a hint. Game designers/educators should select the appropriate type of feedback for each subtask or provide feedback in the form of a report at the end of the ELT cycle. The feedback cannot be effective without the player acting on it. The player must have the choice to take action based on the provided feedback to improve learning and move the learning process to the next step.

The progression of the player influences their encouragement and motivation to continue with the learning process. Game elements can be utilised to show progression to the player, such as a *progression bar* and *unlocking content*. The progression bar is a way of acknowledging the player's progress in the VFTG; then the player can be motivated by good progress or encouraged to work harder if they have made slow progress. Unlocking content such as new tasks, tools, or learning materials as a result of progression motivates the player to continue playing with the aim to access locked content and tools.

4.2.2.2 Task model

Tasks in VFTGs should be shaped based on the understanding of the four stages of ELT and designed to represent each stage and allow the player to go through them all. The task model is the second step of game modelling after defining the structure of the VFTG and selecting suitable game elements based on the matching scheme, and it followed by the evaluation model to design appropriate assessments for designed tasks.

Task-based learning (TBL) is a learning theory that has proved to be effective in teaching languages and uses the task as a unit of analysis where the design of syllabus/instruction broke down into units (tasks) to organize learning. In addition to concentrating on the learning process rather than only on the final product. A task is defined as an activity given to the learner with the goal of achieving the learning outcome. Two main concepts are adopted from TBL: 1) the task forms the unit of designing learning in VFTGs and 2) focusing on the learning process. This conceptual framework considers the task as the unit of learning by organizing the design of VFTGs around tasks; this is going to be clarified in the rest of the section. Also, the conceptual framework models the design of tasks to capture the process in addition to the final product; this is going to be explained in the evaluation model section (4.2.2.3). The expected result is that players who provided authentic and complex goal-oriented tasks would analyse the tasks to a series of interactions based on the required knowledge and skills to solve the task. This results in forming an overlapping network of interactions to achieve the desirable outcomes of the learning process.

The task should be designed with *control of choice* to solve the task in different ways (different interactions) such as using different tools or methods for collecting samples. Also, side-quests could be provided to the player as extra work. This way, each task can be solved by a sequence of interactions and each player can solve the same task with some differentiation in that sequence or do as many extra side-quests as the player wishes. Some of the interactions could be required for more than one task. Based on the analytic method of solving tasks, tasks can be formed as a network (graph) of interactions where each node represents an interaction, and the edge represents the time necessary to perform the interaction.

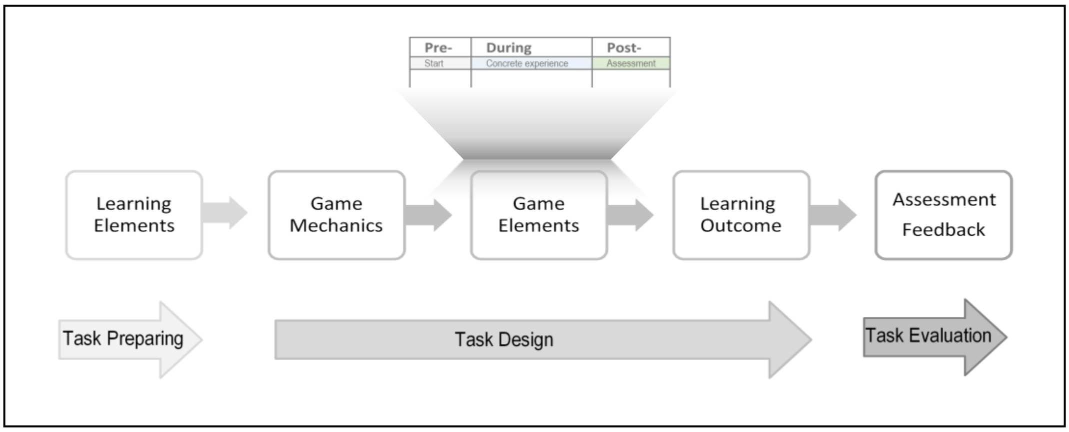


Figure 4.11: Task model.

As a part of this VFTG conceptual framework, where designing learning is constructed around tasks, the task model provides general instructions to help educators/game designers to plan tasks. The task model (Fig. 4.11) consists of three steps: task preparing, task design, and task evaluation. The first step of designing the VFTG is to connect ELT to the internal economy, and the second step is to link the VFTG structure to the FBL structure and define appropriate game elements. The third step involves integrating tasks into the VFTG structure and connecting them to the game elements selected in the previous step. The task model helps to design the pre-task in the *pre-phase* as well the subtasks in the *during* phase.

Task preparing consists of defining some learning elements that are required for designing the learning tasks of VFTGs. These elements are goal, objectives, complexity, and difficulty. The learning goal and objectives have to be specified as a first step where clear objectives help to design suitable game mechanics to ensure motivation and engagement. Learning objectives need to be translated to knowledge concepts that are required to solve tasks or could be gained from solving tasks. This process will be explained in the evaluation model (Section 4.2.2.3).

VFTG tasks have to be planned with opportunities to encourage the learner/player to practise techniques that cannot be implemented elsewhere. Location and time are variables to manipulate to define complexity when designing tasks for VFTGs, as displayed in Fig. 4.12. These two variables have to be adjusted based on the learning goals and objectives of the VFTG, where the combination of different values (short/long and single/multiple) will produce different types of tasks. A single location with a short period of time available allows one task to be performed, such as exploring one level/zone of the VFTG environment, which is suitable for a single player and a linear narrative to be embedded within the learning task.

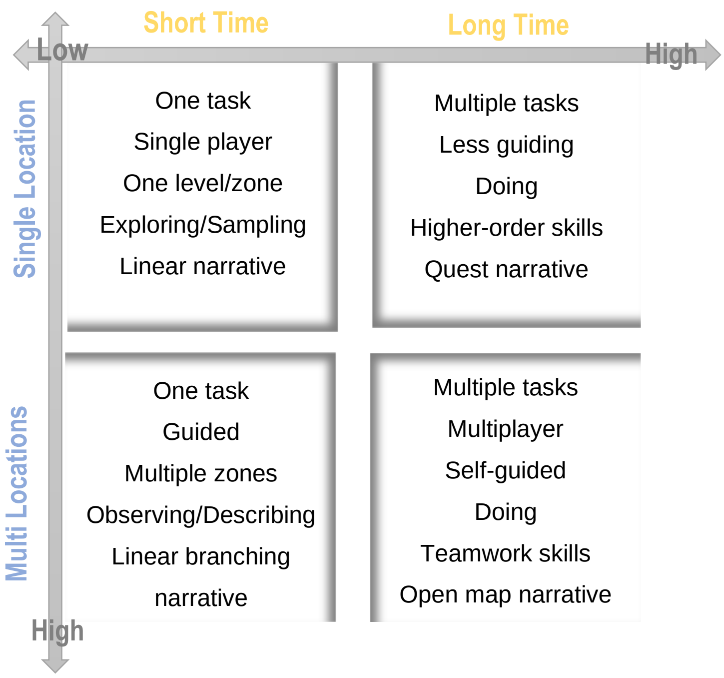


Figure 4.12: Task variables (time & location).

Extending the time in one location permits designing multiple tasks for a single player, who will have more time to explore things and doing (analysing and reasoning) to gain knowledge and skills by him/herself with less guidance. At the same time, a quest narrative can be applied to fit multiple tasks where each cluster of story nodes could represent one of the ELT stages. Extending the location to more than one level or zone with a short period of time available allows the design of a type of task called a "cook's tour", where the player scans and observes rapidly driven from zone to zone collecting a huge amount of information with guidance from the NPC. This type of task could be used to record and reflect on each location, and a linear-branching narrative can be applied where the main narrative is formed in a linear fashion with some side-quest branching to encourage the exploration of more zones.

The last manipulation of time and location applies to multiple locations and a long period of time, which is suitable for designing many tasks to be achieved while progressing from level to level without the need for guidance in a multiplayer VFTG. In addition, in an open map narrative the player can move forwards and backwards between locations exploring the story and the environment. Also, the multiple locations and long period of time scenario provide a suitable VFTG learning environment for repeating the ELT cycle, while designing one ELT cycle can be accomplished with one location and a long period of time. The long period of time provides a VFTG learning environment for doing tasks (hypothesizing, testing, and analysing) beyond exploring and observing (collecting data and describing). The difficulty of the task should be defined and increased regarding the player's progress and level of ability.

The **task design** step involves defining the game mechanics and learning outcomes and linking them to selected game elements based on the matching scheme. The game designer is required to choose appropriate game mechanics (running, searching, role-playing, puzzle solving, constructing, social interaction and tactical strategy) for learning elements defined in the first step, such as applying a searching game mechanic for observing learning objectives, or using constructing mechanics to gain new knowledge. Then, the game mechanic can be connected to appropriate game elements from the matching scheme. For example, multimodal presentation and interaction can enable a searching mechanic, while identity can enhance role-playing and social interaction mechanics. Defining the expected learning outcome for each game mechanic completes the design of a specific learning task. The design of a learning task in the VFTG requires learning elements, game mechanics, game elements, and learning outcomes. The learning outcome is fed into the task evaluation step.

Table 4.2: Example of VFTG game mechanics.

Task type	Game mechanic	Learning outcome	Difficulty
Grasping	Collect specific types of (rock, plant ...) based on the knowledge that is aimed to be taught to the player.	Remembering (define, find, select)	level1
	Locate specific artefacts in a museum/archaeological site or topography in a nature site.		
	Memory mechanic where the player is required to recall earlier (events, knowledge, or collected data) to complete a new task.		
	Repeat pattern, where the player must repeat a sequence of steps in order to memorize or practise a specific procedure such as measuring something in the site or collecting a sample from nature (soil or water).		
	Social interaction by describing a natural event or the use of a tool to another player in a chat.	Understanding (describe, classify, compare)	level2
	Categorize (experiment tools or collected samples) in different containers (boxes or shelves).		
	Compare two game objects related to learning knowledge such as two types of flowers or natural phenomena.		
	Repurpose game object to complete a task such as using straight sticks and stones to create a sundial clock.	Applying (demonstrate, make use of, Interview)	level3
	Puzzle solving mechanic (verbal) that requires demonstration of learned knowledge.		
	Practise conducting an interview with an NPC expert to understand learning concepts or to collect data.		
Transforming	Secret unit deployment (hidden information) mechanic can be used to encourage players to infer the missing information.	Analysing (analyse, discover, distinguish)	level4
	Pattern awareness by challenging the player to recognize a pattern in a science event, in an artist's work, or in the architecture of ancient culture.		
	Information overload mechanic can be applied by presenting the player with clues and information, and the player is expected to analyse and infer valuable knowledge to use it to complete a task.		
	Discuss (a plan or experiment procedure) in a chat with players or NPC.	Evaluating (criticize, prioritize, estimate)	level5
	Time limit mechanic to encourage the player to prioritize tasks based on an educated estimation of the importance of each task.		
	Building/merging mechanic can be used to allow the player to create whatever tool can be used in the VFTG.	Creating (construct, design, formulate)	level6
	Design an experiment or the creation of a tool with players or individually in an electronic note.		
	Collect and combine components of a specific formula to create a medicine.		

Table 4.2 presents examples of game mechanics that can be utilised to design learning tasks based on two types in relation to ELT: grasping experience (CE & AC) and transforming experience (RO & AE). The learning outcomes and difficulty levels can be defined by educators to match the subject of the VFTG. The revised Bloom's Taxonomy along with its levels of difficulty are applied to the examples as a good number of educators are familiar with it.

Story task can be linked to learning task to ensure the element of fun and to keep the player engaged in the VFTG, which ensures the continuation of learning and playing. For instance, the first example of a game mechanic in Table 4.2 is collecting, which can be applied to form a learning task where the player is asked to collect particular plants and to complete the task successfully; he/she needs to recall or search for the distinguished plant's features while observing a natural reserve, which is the environment of the VFTG. This learning task should be designed to match specific learning elements such as a learning objective (this task will encourage players to identify plant features), learning outcome (remembering), and level of difficulty (level1). The task can be connected to the story task such as the threat posed by some insects, increasing the rate of producing this threat for every plant collected by the player. Fig. 4.13 shows an abstract view of connecting learning tasks with story tasks to form the ELT cycle.

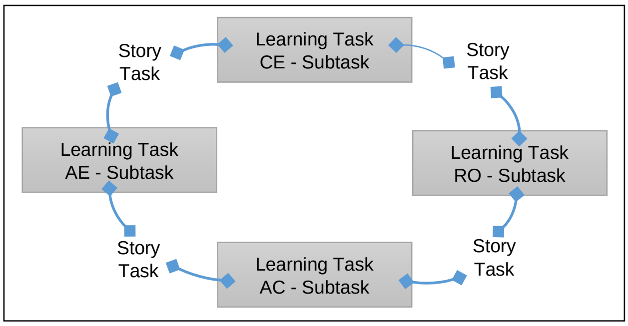


Figure 4.13: Abstract view of learning & story tasks connection.

Task evaluation is the last step in task modelling, where assessment is built upon clear learning elements and blended with the game design elements. Evaluation can be implemented by a traditional method after completing the VFTG or as an in-game assessment. However, some learning tasks can be used for practice without grade assessment. The traditional exam could be an item-based test, presentation, or an essay. However, assessment after completing the VFTG or even physical FT omits evaluating the learning process and focuses only on the learning products. It is difficult to assess the learning process in physical FTs because of the increased number of learners and the limited time of FTs. These limitations can be addressed by in-game assessment, which ranges from something as simple as point scoring to something as advanced as analysing behaviour patterns. Designing in-game assessment could be limited by applying traditional test methods such as multiple-choice questions, which ignore the unique abilities of VFTG environments to store player log data. VFTG environments provide the ability to track learners' interactions and trace their behaviour patterns.

Regardless of the assessment method, assessment criteria should be interweaved into the gameplay and narrative. The selected assessment method has to be suitable for what is to be measured: factual knowledge, procedural knowledge, or high-order skills. For example, assessing collaboration skills could be measured based on the final product and/or the process of collaboration. It could be narrowed to applying a rubric or analysing collaboration in text chat as a process of constructing knowledge or working on a project together. Item-based testing that applies item response theory (IRT), or inputting observation notes into a spreadsheet can be used in VFTGs but should be linked to game mechanics and elements to reduce learning interruption. The assessment has to fit into the VFTG environment and provide feedback that is also intertwined with the VFTG environment and at the same time adheres to the learning elements.

4.2.2.3 Evaluation Model

This framework encourages the design of evaluation to be embedded into the gameplay without interruptions. The evaluation model can be utilised as a building block which can be attached after each stage of ELT or used once at the end of the cycle. The model explains the process of assessment and feedback using game machinations. The three elements of the internal economy are identified and balanced by game machinations. The process of modelling evaluation is broken down into building blocks as follow to help the game designer/educator to develop the VFTG design block by block.

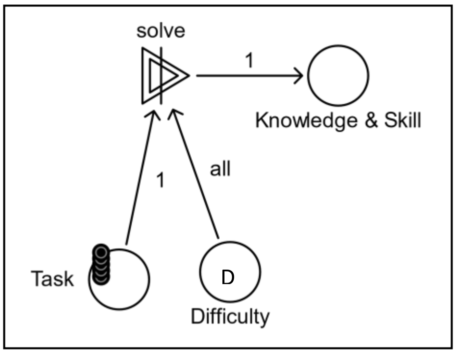


Figure 4.14: Constructing block.

Two resources displayed in Fig. 4.14 start the evaluation process, which are stored in two pools, *Task* and *Difficulty.* The *Task* pool should store the number of tasks; for example, four tasks are represented in the model by four coins, while the difficulty level *D* can be for example equals to 1 or 2 stored in the *Difficulty* pool*.* The player has to solve the task according to its difficulty level by clicking on the interactive *Solve* converter, which is the first internal mechanic applied in this evaluation model. The *Solve* converter will receive one task and all the content of the *Difficulty pool* through the resource connections that connect the *Task* and *Difficulty* pools to the *Solve* converter. The labels of the *Task* and *Difficulty* pools show the conversion rate as 1 and all, respectively. The *Solve* converter will construct knowledge and skill resources and store them in the *Knowledge & Skill* pool. This simple step represents the knowledge constructing block.

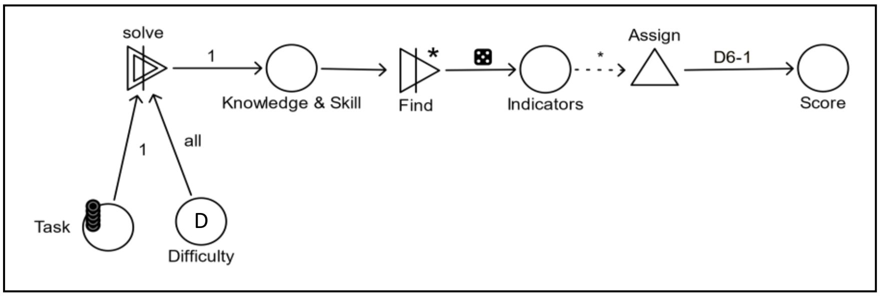


Figure 4.15: Finding indicators Block.

When the *Knowledge & Skill* resource becomes available as a result of the construction process, the automatic *Find* converter will work to convert knowledge and skill to indicators and store them in the *Indicators* pool, as shown in Fig. 4.15. The rate of the *Find* converter is indicated by a dice in the label of the resource connection between the *Knowledge & Skill* pool and the *Find* converter to represent the randomness of finding indicators which should differ from player to player based on ability. Finding indicators will trigger the *Assign* source via a trigger state connection (dotted arrow labelled with an asterisk *) to determine the score for that task; the rate for assigning a score could vary from 0 to any maximum score, as shown by the label of the resource connection (D6 – 1). D6 means a dice with six sides, representing the variety of scores and the -1 brings the chance of scoring zero as there is a possibility of not successfully completing the task. The score generated by the *Assign* source should be stored in the *Score* pool.

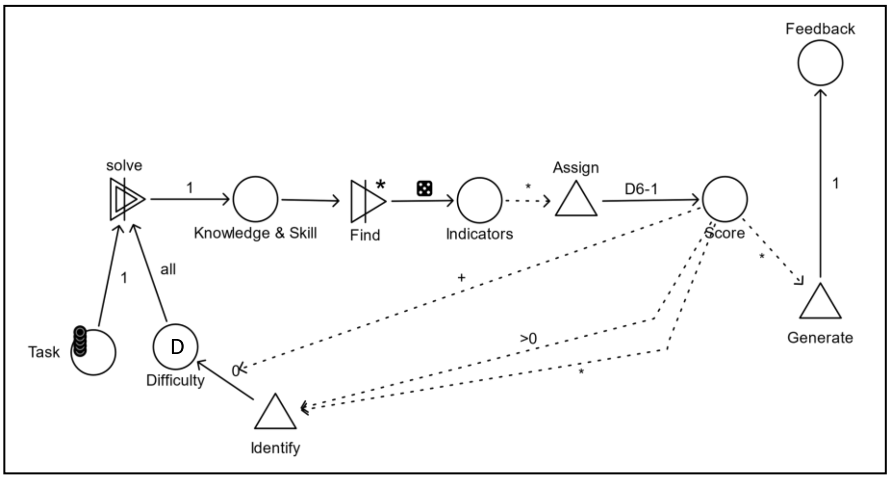


Figure 4.16: Difficulty balance block.

Storing scores in the *Score* pool will trigger the *Generate* source through a trigger state connection to provide feedback and store it in the *Feedback* pool, as displayed in Fig. 4.16. In addition to triggering the *Generate* source, storing scores in the *Score* pool will activate the *Identify* source to identify the difficulty level for next task based on the condition of an activator state connection (dotted arrow labelled with the condition > 0). Also, the *Identify* source will be triggered through a trigger state connection. Both the activator and trigger state connections will then allow the difficulty level to be increased if the score in the *Score* pool is greater than zero.

The difficulty level of the next task will be increased based on the change in resources of the *Score* pool as represented by (+) on the label of the label modifier state connection (dotted arrow with + from *Score* pool). This will affect the rate of the *Identify* source each time a change occurs in the resources of the *Score* pool. If the score in the *Score* pool is zero, the difficulty level will reset to zero, which could be considered as a default level of difficulty where the previous task should be repeated.

Fig. 4.17 shows that the player has the choice to reflect on the feedback by clicking on the interactive *Reflect* converter to take action based on that feedback to reduce the gap between the expected solution and the player's solution. The acting on the provided feedback is represented as a resource and stored in the *Action* pool.

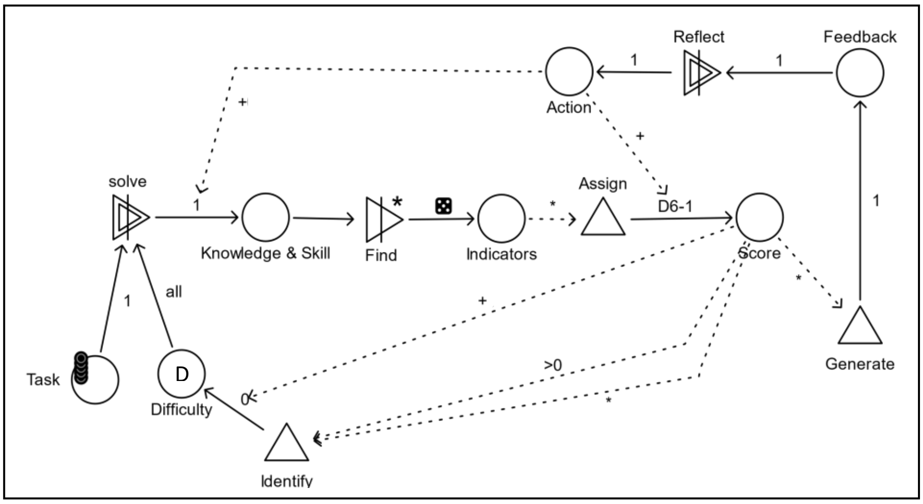


Figure 4.17: Feedback action block.

As the *Action* pool has new resources, the rate of constructing knowledge and skill for the next task will be boosted by increasing the rate of the resource connection of the *Solve* source by some value through a label modifier state connection (dotted arrow with +). This shows the expected improvement of the learner's performance after acting on the provided feedback. Acting on the feedback also should affect the score of the next task by increasing the rate of resource connection of the *Assign* source by some value through a label modifier state connection (dotted arrow with +) between the *Action* pool and the label of the *Assign* source. The mode of both the *Solve* converter and *Reflect* source is interactive (double outline) as both require player action.

Three trigger state connections will work when the player clicks the interactive *Solve* converter to solve the next task, as displayed in Fig. 4.18 which shows the complete evaluation model. The first trigger state connection will trigger the *Assessment Record* pool to pull all resources of the previous task from the *Indicators* pool. The second trigger state connection will trigger the *Total Score* pool to pull all scores of the previous task from the *Score* pool, which will set the label of the resource connection of the *Identify* source back to zero to not accumulate an increase in the *Difficulty* source. The third and last trigger state connection will trigger the *Report* pool to pull all the feedback generated for the previous task as the player lost the chance of reflecting on it when a new task started to be solved. If the feedback is moved to the *Report* pool, this means that the player did not reflect and act, which leads to a decrease in the rate of the resource connection of the *Assign* source by some value, showing an effect on learning performance as there will be a limited impact on reducing the learning gap.

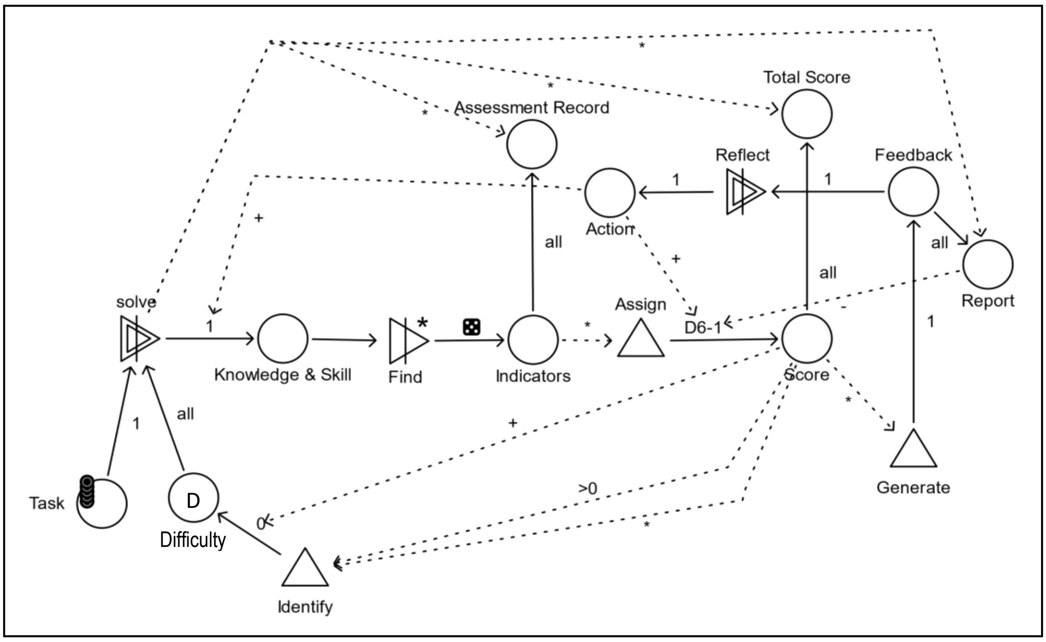


Figure 4.18: Evaluation model.

At two points in the evaluation model, the resources are balanced so as to not overwhelm the assessment process with scores or difficulty levels. First, to balance the assignment of scores, the rate of the *Assign* source is increased as the player reflected on feedback and took action. In contrast, the rate of the *Assign* source is decreased by moving feedback to the *Report* pool at the start of the next task to create a balance in player scores. Second, to balance the increase in difficulty level, assigning and storing scores in the *Score* pool will lead to an increase in the level of difficulty for the next task, while moving scores from the *Score* pool to the *Total Score* pool will set the difficulty level to zero.

The indicators, which are important resources produced in the evaluation model, need to be specified based on particular interactions in the VFTG as evidence of constructing knowledge and improving skills. Usually, learning tasks are very specific to an individual game or particular subject. However, learning tasks in the VFTG can be defined based on an expected series of interactions to be performed by the player. Defining the learning tasks regarding to expected interactions results in generalising and standardising the assessment regardless of the subjects or audience. Then the learning outcomes can be inferred from the player's interaction. The evaluation model forms interactions in VFTGs in a way to capture indicators of learning performance. This evaluation model introduces three types of indicators for assessment in VFTGs: knowledge indicator, FB indicator, and skill indicator. The knowledge indicator distinguishes players' interactions that reveal gained/employed subject knowledge.

The FB indicator detects players' interactions that show acquired FB knowledge. The skill indicator identifies players' interactions that point out developed skills.

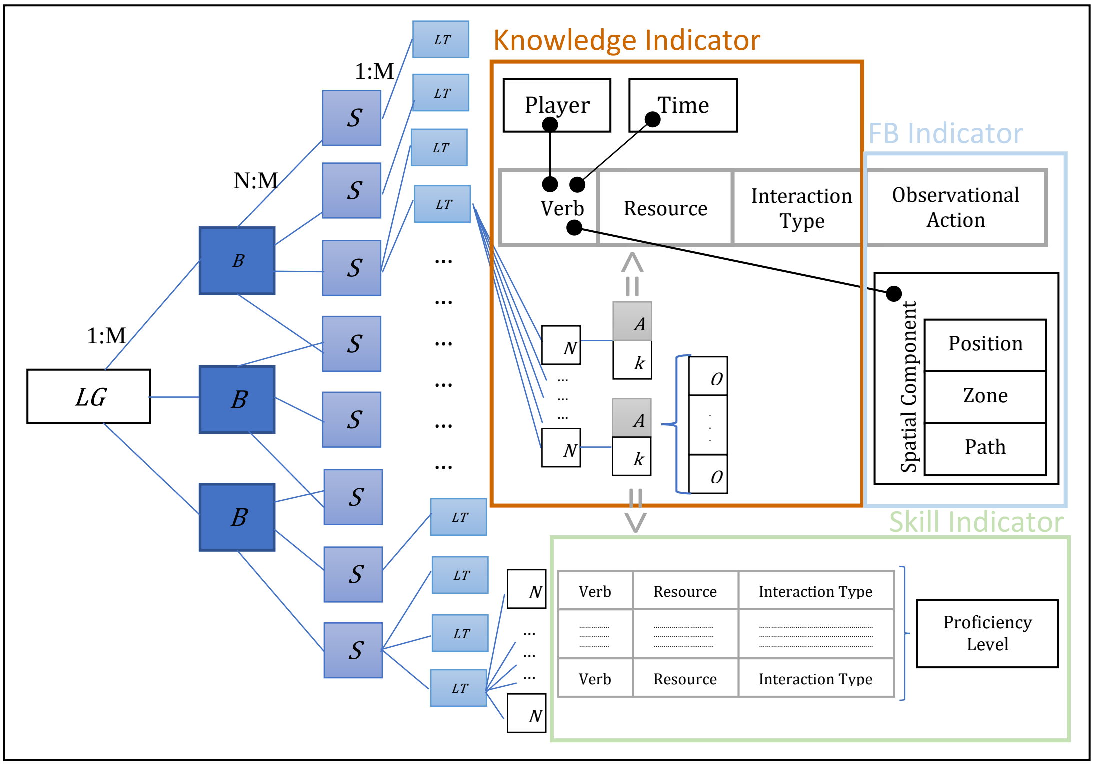


Figure 4.19: Illustration of identifying indicators.

Identifying indicators facilitates the task evaluation process, which provides a way to connect game design elements with learning elements as the designing process developed. Fig. 4.19 demonstrates the process, which starts with identifying the learning goal in connection with actions and their related attributes. Table 4.3 presents the abbreviations and definitions of both learning and game elements utilised in the process of identifying indicators. The process displayed in Fig. 4.19 is explained in detail below.

Table 4.3: Learning and game elements used to identify indicators.

Abbreviation	Component	Definition
LG	Learning goal	A description of the overall purpose of the VFTG.
B	Learning objective	A description of what educators intend to teach their learners (knowledge, skill, or behaviours) by the end of the VFTG in order to meet the learning goal.
KU	Knowledge unit	A broad knowledge topic which consists of a set of smaller knowledge concepts.
K	Knowledge concept	Individual knowledge concepts that can combine to form KU.
PK	Pre-requisite	A graph that defines the KU dependencies.
N	Interaction	A gameplay interaction between the player and game object (O) with the aim of completing a specific task.

Table 4.3: Learning and game elements used to identify indicators (Continued).

Abbreviation	Component	Definition
A	Action	Action verb to perform as a part of the gameplay interaction.
OA	Observational action	An action performed to solve *LTs* that require the player to have the ability to visually recognize features or events in the virtual field.
O	Object	Game object that is subject to an action by the player.
TN	Type of interactions	An element of a set that defines the type of interaction as one of three *{Social, Environmental, Learning content}*.
LT	Learning task	A gameplay task designed to help the player to achieve a learning objective.
AR	Assessment result	The total result of assessing the *LT* based on all interactions performed by the player to complete the *LT*.
P	Player	An indicator to identify the player who made the interaction in a multiplayer VFTG.
CT	Time	An indicator of the time required to perform a *LT*.
SI	Spatial indicator	An indicator of spatial information about the player's movement in the VFTG environment and can be about the position, area, or path.
SK	Skill indicator	Collection of interactions of each required step to perform a skill along with proficiency level.
ARS	Assessment result of skill	The sum of assessment results of LTs that form a skill indicator.
PLS	Proficiency levels	Illustration of how a particular skill is performed at different levels of proficiency.
CD	Conditions	A binary vector that reflects conditions of proficiency levels and presents the fulfilled condition as 1 and the rest as zeros.

VFTGs are designed based on a learning goal, which is a description of the overall purpose of the learning in general. Each VFTG can have one learning goal explaining what players will gain from the VFTG broadly, and the VFTG consists of a set of learning objectives derived from the learning goal, which should be already identified by the task model. Learning objectives describe what educators intend to teach their learners (knowledge, skill, or behaviours) by the end of the VFTG in order to meet the learning goal. Game designers/educators can define the learning goal (LG) as a composite of several learning objectives with different weights of importance such that:

$$LG = [w_1 \; .... w_i \; .... w_{|B|}] \, [B_1 \; .... B_i \; .... B_{|B|}]^T \quad for\ 1 \leq i \leq |\, B| \qquad (4.1)$$

Where w_i is a weight to signify the importance of the learning objective (B_i). The value of w_i ranges from 0 to 1 and $\sum w_i$ equals to 1, while $[\;]^T$ is a transpose function and $|B|$ is the cardinality of B. The educator/game designer can discuss the importance of each learning objective, which should be included in the VFTG first, and which are considered less important so that they can be skipped or saved for later in the VFTG. Each learning objective can be designed to be accomplished in one or more scenes/levels (S) and each scene can support one or more learning objectives. The relationship between B and S is many-to-many, while it is a one-to-many relationship between LG and B.

The knowledge indicator gathers evidence of specific subject knowledge, such as natural hazards in geography FTs or historical events in a museum trip. A list of knowledge units (KUs) has to be identified where each knowledge unit is a broad knowledge topic consisting of a set of smaller knowledge concepts (K). All knowledge concepts of a specific knowledge unit have the same level of importance. For a player to be competent in a particular knowledge unit he/she has to be competent in all the knowledge concepts forming that knowledge unit. The knowledge unit is defined as follows:

$$KU = \{K_{1,}\ldots, K_{i,}\ldots K_{|KU|}\} \;\; for\ 1 \leq i \leq |KU| \tag{4.2}$$

Where K_i is a knowledge concept and $|KU|$ is the cardinality of KU. For example, *Place Knowledge* is a KU composite of two knowledge concepts *{human geography of the Africa region, physical geography of Africa }*, and $|KU|$ = 2. *Place Knowledge* could consist of more than two knowledge concepts but these two are the only ones required for that specific VFTG.

The pre-requisite KU is modelled as a weighted-directed graph $PK = (VR, E)$, defining the KU dependency where VR is a set of n vertices of knowledge units $\{KU_{1,}\ldots, KU_{i,}\ldots KU_n\}$, E is a set of edges connecting pairs of KUs, donating the weight of edge (i, j) by w_{ij}, and a matrix $Q(PK) = [q_{ij}]$ such that:

$$q_{ij} = \begin{cases} w_{ij} & if\ \ i \neq j\ and\ (i,j) \in E \\ -1 & if\ \ i = j\ and\ (i,j) \in E \\ 0 & if\ \ i \neq j\ and\ (i,j) \notin E \end{cases} \tag{4.3}$$

Where w_{ij} shows the level of dependencies between KUs, where 0 implies no dependency between the two nodes of KUs, and -1 shows that should not be any dependency between KU and itself.

The game designer can prioritize the KU with a greater weight to involve that KU in an earlier level/zone before the lesser one. For example, *Ancient Rome* is a KU that has three pre-requisite KUs *{Early History of Rome, Roman Republic, Rome Culture}* with different levels of dependencies: 0.3, 0.3, 0.4, respectively. Therefore, *Rome Culture* should be the first KU to be designed and introduced in the first scene/level of the VFTG. The clear declaration of KUs and their pre-requisites will help in defining interaction and the order of scenes/levels.

Interaction (N) is a gameplay interaction performed by the player in order to complete a specific task. Each task can be completed by performing one or more interactions. Tasks in VFTGs can be categorized into two types: practice or assessment tasks. practice tasks are planned to help players to gain the intended knowledge or to develop the intended skills by performing gameplay Ns, while assessment tasks evaluate the expected knowledge or skill by comparing the player's collected indicators of Ns with the predefined values of indicators of the intended knowledge. This way, a learning gap can be found, and proper feedback can be provided. The indicators of practice tasks can be used for adapting these tasks to the player's ability level but not for calculating scores. Interaction (N) is defined based on an action (A) performed by the player on one object (O) or more by applying a specific K as follows:

$$N = K * A + [O_{1,} \dots, O_{i,} \dots O_{|O|}] \;\; for\ 1 \leq i \leq |O| \qquad (4.4)$$

Where K is considered as constant and A as a variable and together they combine to O that will receive the action. The same K can be employed to perform different As and one or more objects may receive the same A. A is defined by a function $M(\,)$ to map a single required resource and type of interaction to an action verb such that:

$$A = M(Verb, R, I_i), \;\; I_i \in TN \qquad (4.5)$$

Where $Verb$ is an action verb such as collect or measure, R is a resource that could be required to do the action such as a tool to collect a sample of a river's water, and I_i is an element of a set of type of interactions (TN) which includes three elements as follows:

$$TN = \{Social, Environmental, Learning\ content\}$$

For each N, the game designer can specify two attributes: level of difficulty and criteria (quantity or quality). For example, the action should be performed at least on two objects over a specific period of time. The learning task (LT) is formulated as a sequence of interactions as follows:

$$LT = [w_1 \;\dots . w_i \;\dots . w_{|LT|}]\, [N_1 \;\dots . N_i \;\dots . N_{|LT|}]^T \; for\ 1 \leq i \leq |LT| \qquad (4.6)$$

Where w_i is a weight to signify the importance of interaction N_i and ranges from 0 to 1 and Σw_i equals to 1. While $[\ ]^T$ is a transpose function and $|LT|$ is the cardinality of LT. Some interactions could be more important and so should be more weighted and then graded with higher scores, or some interactions are equivalent, but a player needs to perform all of them to complete the LT.

The final step is assessing the indicators regarding to associated learning outcomes. The assessment method should be selected to be suitable for the VLE such as giving XP for the interactions that a player performed or classifying the player's performance as novice or expert via a clustering mechanism. The level of difficulty and the criteria of interaction play a role in the assessment as a higher level of difficulty results in a higher score and if a player's interactions meet all the criteria, then this will result in a better score. The assessment result (AR) of LT equals the assessment results of each interaction that forms the LT multiplied by its weight as follows:

$$AR = [w_1 \dots w_i \dots w_{|LT|}][S_1(N_1) \dots S_i(N_i) \dots S_{|LT|}(N_{|LT|})]^T for\ 1 \leq i \leq |LT| \quad (4.7)$$

Where $S_i()$ is an assessment method which should be defined by the game designer to be applicable to N_i. The weights are the same as the weights coupled with Ns that compose the LT while $[\ ]^T$ is a transpose function and $|LT|$ is the number of Ns that make up the LT.

There are two more indicators which are optional: a player indicator to identify the player (P) in multiplayer VFTGs, and a time indicator if time is an important component of interaction and influences the assessment. The time indicator defines two time points, start (Ts) and end (Te), where they can be collected for the N and then the completion time (CT) can be calculated as the difference between the end point and start point.

$$CT = Te - Ts \quad (4.8)$$

As LTs could require performing more than one N, the completion time (CT) of the LT is computed by summing the completion time of each performed N.

$$CT_{LT} = \sum_{i=1}^{c} CT_i = CT_1 + CT_2 + \dots + CT_c \quad (4.9)$$

Where CT_{LT} is the completion time of one LT, and c equals the number of interactions performed to complete the LT. CT_{LT} is considered as the temporal length of completing an LT, while the number of performed Ns might vary in some cases when the player is given control of their choice to complete the LT by a different set of Ns and provided side-quests that could enhance the final achievement of the LT. Therefore, the topological length (c) can be calculated as the number of Ns performed to complete a particular LT, which may affect the assessment result in some way such as more Ns means higher points or better health.

$$|CT_{LT}| = c \tag{4.10}$$

FB indicators combine observational and spatial indicators. The observational indicator captures the player's ability to visually recognize features in the virtual field. Observation is an important and distinguished skill of FBL. In VFTGs, the player can develop observational skills by prompting identifying features or events in the virtual field, where the player could be asked to describe, make a list, or solve any other learning task (LT) based on visually recognized features or events. These are observational actions (OAs), such as *Describe* as OA can be designed to be part of a chat discussion between two players or a written task in electronic notes, while *List* as OA can be designed as collecting/shooting objects with a specific feature. Observational indicators can be distinguished and stored by adding a binary variable to $M(\,)$, where a value of 1 means it is an OA and a value of 0 is given otherwise. A can be redefined as follows:

$$A = M(Verb, R, I_i, OA),\ I_i \in TN \ and\ OA \in \{0,1\} \tag{4.11}$$

The spatial indicator (SI) provides information about the player's movement in the VFTG environment and can be about the position, area, or path. The spatial indicator can be essential to evaluate learning tasks in VFTGs because the player's position or path affects the exploration and discovery of the VFTG environment, which is reflected in gaining knowledge and developing skills and will influence solving learning tasks. Evaluation of regular GBL would not consider observational and spatial indicators as very important for assessing players' performance in general. Spatial indicators can be collected and stored regarding different types of spatial components. The position type requires capturing the (X,Y,Z) vector of the player's position, and the area type requires a single variable that indicates the name or code number of the area. The movement path of the player is a series of player positions captured at certain intervals.

$$SI = \begin{cases} (X,Y,Z) & if\ SI\ is\ a\ position \\ Code & if\ SI\ is\ an\ area \\ [(X_1,Y_1,Z_1),(X_2,Y_2,Z_2),\ldots] & if\ SI\ is\ a\ path \end{cases} \tag{4.12}$$

The skill indicator (SK) captures the development of high-order skills required in FTs such as decision-making and teamwork skills. The skill indicator should be defined based on a high level of abstraction to be suitable for evaluating different skills. A skill is measured based on steps that are supposed to be implemented by the player to show the level of proficiency. Each step represents one LT while the steps can be defined based on a specific theory or model. For example, the process and steps of the decision-making skill are characterized by different models/theories: Intuitive (Sauter, 1999), Game theory (Xiao, Zeng, Allen, Rosen, & Mistree, 2005), Bounded rationality (Camerer, 1998), and Rational (Heracleous, 1994). Educators/game designers need to select the suitable skill development model or create a new one and utilise the model to map the steps of performing the skill to LTs, where the skill indicator is represented as a vector of learning tasks:

$$SK = \begin{bmatrix} LT_1 \\ \vdots \\ LT_{|SK|} \end{bmatrix} for\ 1 \le i \le |SK| \qquad (4.13)$$

Where each row of the vector is equivalent to Eq. 4.6, and thus the assessment result can be computed by Eq. 4.7 for each LT_i when applicable assessment methods are applied. $|SK|$ is the number of steps required to perform a skill. Skill evaluation follows one of two methods: assessing the final product of performing a specific skill or assessing each step individually and then summing the assessment result of all steps as follows:

$$ARS = \sum_{i=1}^{n} AR_i = AR_1 + AR_2 + \ldots + AR_n \qquad (4.14)$$

Where ARS is the total assessment result of skill performance, AR_i is the assessment result of a single LT presented in Eq. 4.7, and n equals to $|\mathrm{SM}|$, which is the number of steps required to perform a skill.

Skill acquisition is usually labelled based on levels of proficiency such as novice or expert, following clear criteria for each level. Again, educators/game designers can select their preferred proficiency model or define a new one. Proficiency levels (PLS) can be set up as a row vector containing the levels as follows:

$$PLS = [L_1, \ldots, L_i, \ldots L_{|PLS|}]\ for\ 1 \le i \le |PLS| \qquad (4.15)$$

Where L_i is one level of proficiency and $|PLS|$ is the cardinality of PLS. For example, the PLS of the Dreyfus model of skill acquisition (Dreyfus & Dreyfus, 1980) consists of five levels:

$$PLS = [Novice, Advanced\ Beginner, Competent, Proficient, Expert]$$

The player's performance of a skill can be labelled based on the selected proficiency levels, where assigning a level depends on a specific condition to be fulfilled. For each level of proficiency, a condition should be defined. For example, the conditions of the Dreyfus model could be specified as follows:

1. No final product was produced, which suggests that the player is a Novice.
2. Guidance or a hint is required by a player considered as an Advanced Beginner.
3. The final product is achieved after more than one trial implies that the player is Competent.
4. The player shows the ability to use resources to professionally perform the skill, which indicates that the player is Proficient.
5. The ability to perform the skill successfully under pressure suggests that the player is an Expert.

These conditions can be presented as a binary vector as follows:

$$CD = [D_1, \dots, D_i, \dots D_{PLS}]\ for\ 1 \le i \le |PLS|\ and\ D_i \in \{0,1\} \qquad (4.16)$$

Where $|PLS|$ is the number of proficiency levels identified in the PLS and the value of D_i equals one if the condition is fulfilled and equals zero otherwise. Therefore, the labelling of proficiency is calculated as a multiplication of the conditions vector by the proficiency level vector in Eq. 4.15, where all unmet levels will be eliminated by the multiplication to zero value of unsatisfied conditions:

$$PL = CD * PLS^T \qquad (4.17)$$

Where PL is one value of PLS and $[\ ]^T$ is a transpose function.

Feedback is an essential component of learning and assessment and can be generated by a function:

$$FB = h(result, type, mode) \qquad (4.18)$$

Where $result$ can be the AR, which is the assessment result of a complete LT or ARS, which is the assessment result of the skill indicator. The $type$ of feedback can be specified by the educator/game designer as a set of different options such as knowledge of result (KR), knowledge of correct response (KCR), and elaborated feedback (EF). The same concept is applied to $mode$, which can be defined as a set of different alternatives such as visual, auditory, and haptic.

4.2.2.4 Skill Model

Skills are essential to FBL and educators/game designers need to design and develop VFTGs to improve the required skills for FTs such as problem solving, decision-making, and teamworking within the limitless possibilities of virtual field trips in game environments. A skill task is a composite of learning tasks as steps defined based on a specific theory or model. Defining and explaining skills as an internal economy mechanic supports designing skill tasks in VFTGs. This research presents two skills (decision-making and teamworking) as internal economy mechanics by identifying the three main components (resources, internal mechanic, and feedback loop). Game designers/educators can adapt this research to define different skills as the internal economy mechanic.

Decision-making Skills

Decision-making skills are considered as the capability to recognize needs, assess comparable alternatives, and finalize to select the best option. Decision-making skill builds on the investigation, gathering and analysing of data with aim of reaching the correct decision for desirable final result. Usually, decision-making skills are developed for individuals such as leaders and managers, and as mentioned before there are several models of decision-making skills: Intuitive (Sauter, 1999), Game theory (Xiao et al., 2005), Bounded rationality (Camerer, 1998), and Rational (Heracleous, 1994). However, working with teams in the 21st century leads to making many choices that will affect other team members which result in developing models and frameworks for decision-making skill in teams. some elements can affect the modelling of decision-making skills such as risk and uncertainty (R. Gregory & Long, 2009). This research will use simple steps to model the decision-making skill for individuals to create tasks that are suitable for school students as follows:

1. Identify the problem/situation.
2. Distinguish a list of possible options.
3. Weigh options accurately and determine priorities.
4. Select the decision.
5. Implement the selected decision.
6. Evaluate the effect of the selected decision and keep reference for future decision-making situations.

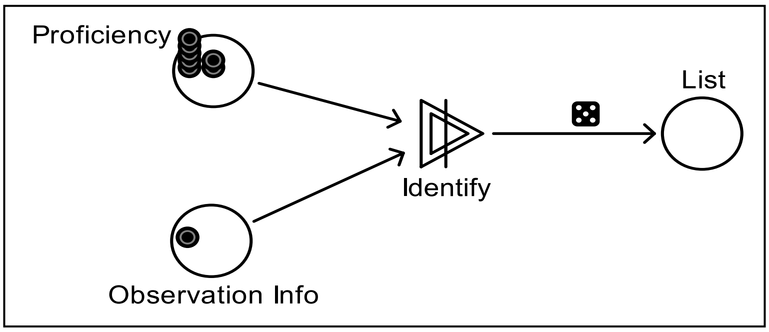


Figure 4.20: Identifying the situation and the list of possible options.

The first two steps of modelling the process of making decisions depend on proficiency level, which defines the level of player's performance of the skill, and the observation information, which results from the first stage of ELT (CE) and can be considered as *level1* resource. The *Identify* interactive converter presented in Fig. 4.20 uses these two resources (*Proficiency* and *Observation Info) to* identify the problem/situation and distinguish the list of possible options and then store it in *List* pool. This step exemplifies the second stage (RO) of ELT as players reflect on their experience from the first stage including observation info and the *List* resource represents *Level2* resource. The number of possible options identified by the player differs based on the situation in hand, for that the production of *Identify* converter is represented as dice to show the random differentiation.

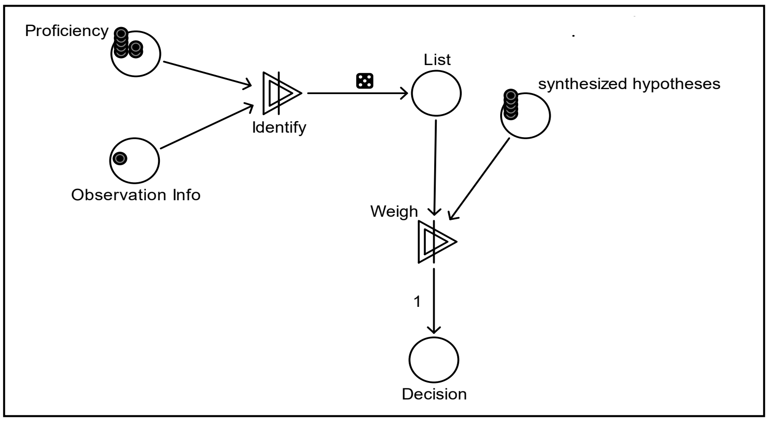


Figure 4.21: Weigh the list of options.

The following step is weighing the list of options to select the best decision. The step is represented as *Weigh* interactive converter as displayed in Fig. 4.21 which works on the options stored in *List* pool to compare them based on *Synthesizing hypotheses* resource which is *level3* resource. The player is expected to develop some hypotheses (*level3* resource) during the third stage of ELT (AC) and can be utilised to make the decision on which option is the best one. The *Weigh* converter uses two resources:

List and *Synthesizing hypotheses* to produce only one decision which is represented as a rate of production equals 1 on the converter resource connection. The selected decision will be stored in *Decision* pool.

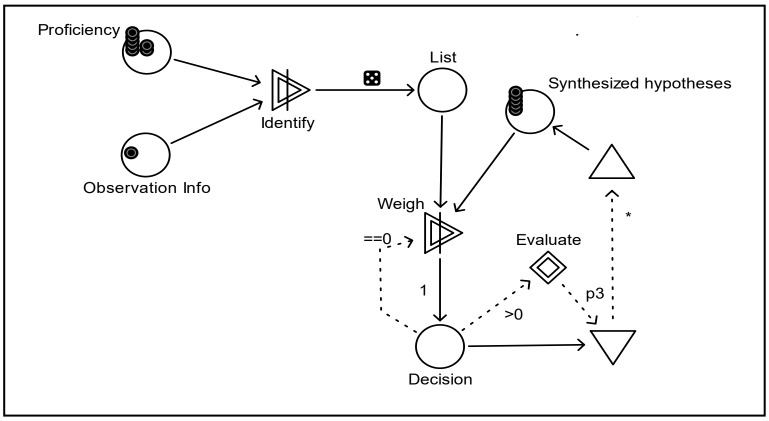


Figure 4.22: Evaluate the selected decision.

The player may have a chance to choose to evaluate the selected decision before implementing it if there is more time or a discussion with team members in case of working in teams. Storing the selected decision in *Decision* pool will activate the interactive *Evaluate* gate by the activator state connection and the condition specified on the label as (0>) as displayed in Fig. 4.22. The player has the chance to evaluate the selected decision by clicking the *Evaluate* gate. When clicked, it has a p3 percentage chance that the player found the selected decision is not the best one which will trigger a drain that drains the selected decision.

Each time the selected decision is drained as found not to be the best one, the *Weigh* converter will be activated again to re-weighing the list of possible options. This is achieved by activator state connection with the condition (==0) between the *Decision* pool and *Weigh* converter. P3 should represent a high percentage at the start such as 60% to reflect an inexperienced player who needs more thinking to take the right decision, and this percentage would be decreased with practice which is explained in the following step. The *Evaluate* gate exemplifies the block of third stage of ELT where it involves the same concept as synthesizing and reflecting" which is similar to "weighing and evaluating" until reaching the final decision. If the selected decision was drained that will trigger a source via a trigger state connection to produce a new hypothesis to be stored in *Synthesized Hypotheses* pool and used for re-weighing the rest of options in *List* pool.

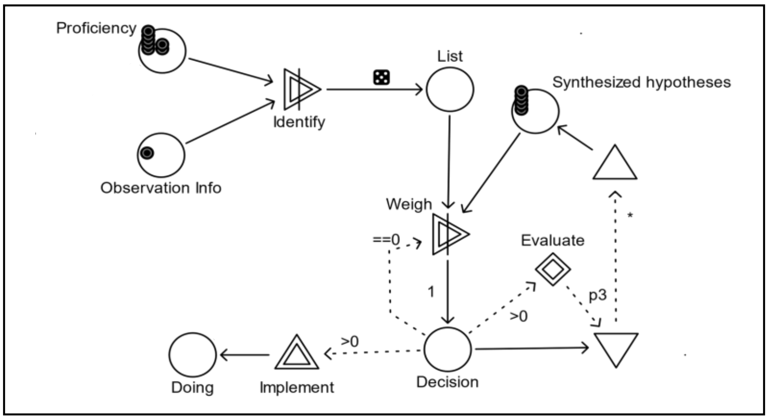


Figure 4.23: Implementation of the selected decision.

The Fig. 4.23 shows the implementation step where the interactive *Implement* source will be activated as the selected decision stored in *Decision* pool via the activator state connection and the condition specified on the label as (0>). The implemented decision will be stored in *Doing* pool and both *Implement* source and *Doing* pool mimic the final stage (AE) of ELT cycle. The evaluation model (Section 4.2.2.3) can be applied to the *Doing* resource.

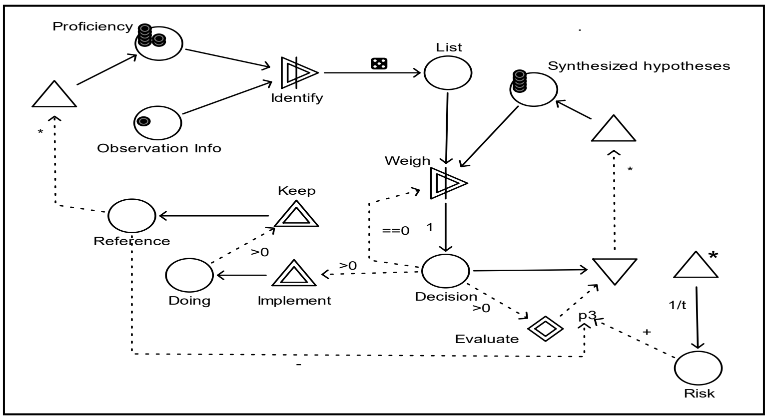


Figure 4.24: Balancing the player’s skill level and task difficulty.

Fig. 4.24 shows the player has the choice of deciding to keep some references by clicking *Keep* source to store references in *Reference* pool. The *Keep* source is activated by storing a resource in *Doing* pool from the previous step. This process could be considered as some sort of self-feedback to improve his/her decision-making skill. Storing references in *Reference* pool is supposed to increase the level of proficiency for the next task by triggering a source via trigger state connection to produce proficiency resources and store them in *Proficiency* pool.

The development of proficiency level of decision-making skills and the difficulty of the task should be balanced. Keeping references from previous decision-making skill tasks should improve the player's ability by reducing the chance of selecting the wrong decision. This is achieved by a label modifier, where the chance p3 of not selecting the best decision will be decreased (-) by some percentage automatically each time a resource is stored in the Reference pool. On the other hand, the chance p3 of not selecting the best decision will be increased (+) by some percentage when a new risk is presented in the environment such as running out of time via a label modifier.

The game designers/educators can use this decision-making skill model as a reference and adjust some elements such as using uncertainty instead of risk to add balancing to the increase of task difficulty or to add teamwork skills to the evaluate process.

Teamwork Skills

The social setting can define a direct experience with aim of advancing the efficiency of learning. Working with other players to construct knowledge or observing other players' experiences is a fundamental part of the learning process according to social learning theory (Bandura & Walters, 1977) The social interaction inside a VFTG along with gaining teamwork skills could lead to players' appreciations of community of practice (Wenger, 1998). The social interaction is explained and utilised by the matching scheme tool and the evaluation model. Yet, the social setting is considered while modelling teamwork skills. Social aspects have been already considered for FBL (Mogk & Goodwin, 2012; Streule & Craig, 2016). Social learning theories distinguish two sides of learning: learning of meaning and learning of practice. The learning of meaning results from giving meaning to learning by placing learners in a real-world situation which is what this conceptual framework models via designing the experience (ELT modelling), constructing the VFTG structure along with suitable game elements (matching scheme), defining the important elements of designing learning tasks such as complexity variables – time and location (task model), defining interactions to solve tasks (evaluation model), and defining VFTGs world variables (world modelling). While the learning of practice results from doing and also this framework models this via the three links. However, social learning theories aim to achieve these two sides of learning in social setting and for that the teamwork skill model is constructed to help game designers/educators to utilise it when they considering social learning.

To model teamwork skills, the basics of teamwork will be discussed. Teamwork skills involve the ability to reach a common goal by collaborating (interaction and sharing) with team members. Examples of teams are a project team, research team, or disaster response team. Teamwork skills are competencies that team members have to retain in order to perform effective teamwork. These competencies include factual information required to perform the task, and attitudes towards collaboration and communication. Teamwork skills is a process of forming a team, defining the roles of each member, identifying and gaining the necessary factual knowledge and attitudes, and finally practising teamwork skills.

In the workplace/field, teams usually represent the core of the organization, most likely because the complexity of tasks cannot be achieved by an individual person. Group tasks are considered as teamwork and are assumed to develop the necessary skills for teamwork. However, in the task attention is paid to the outcomes regardless of the process, which leads to each member working on part of the task individually and at the end the team members combine their parts to produce the final product. That means the team members do not work together, which affects the quality of the final work and does not improve the teamwork skills of those members. VFTGs can be designed to encourage players to work together (collaboration) and individually (competition) and at the same time capture the process of achieving teamwork skills alongside the final product.

For designing learning tasks of teamwork skills in VFTGs, two points are considered:

- Defining teamworking scheme: competition or collaboration.
- Defining level of scheme: individually or in a team.

A team can be formed by allowing the player to pick his/her own team or by creating an algorithm to assign members to teams. The team formation method is out of the scope of this research; however, there is one point worth mentioning regarding the team formation method which would impact players' *control of choice* in VFTGs. Assigning players to a specific team or role would limit the player's *control of choice* in terms of specifically selecting an avatar where the avatar is linked to role-playing and identity. This limitation of choice would affect motivation and immersion, but it would benefit the player's learning when they are assigned to a more suitable team rather than a friendship group, based perhaps on learning styles or functional responsibilities of tasks.

Teamwork produces a product where team members are expected to work together collaboratively (team task); in addition, each member might be asked to participate with his/her individual effort (individual task). Therefore, the designing of learning tasks and evaluation process should consider two components: team effort and individual effort. Working in teams on physical field trips might involve interactions with other teams, in which case the designing and evaluation of learning tasks should consist of two levels: collaboration and competition between teams. Intra- and inter-collaboration/competition is supported by the assumption of learning as a more social act than an individual process. *Inter-collaboration* between two teams could be accomplished by sharing information to help the other team to complete their own task.

The first step in modelling teamwork skills is the production of one team, as shown in Fig. 4.25. The teamwork skill tasks assume that players start with some level of proficiency, that reflects the previous skill of the player, which counts as a resource and is stored in the *Proficiency* pool. In addition to the *Proficiency* resources, modelling teamwork skills could use other resources such as *level1* from the first stage of ELT as it is used in decision-making skills. Another example of utilising different resources in teamwork skills modelling is *Synthesizing hypotheses* if the decision-making task is combined with teamwork skills.

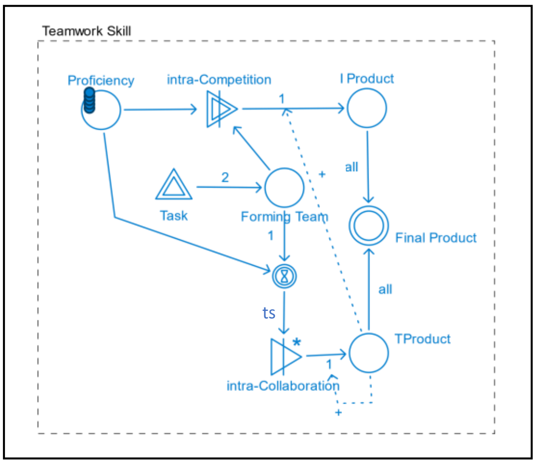


Figure 4.25: Individual and team production.

Forming a team and setting up individual/team tasks are considered as an implicit process, which is out of the scope of this research. This step commences by clicking on the interactive *Task* source to form the team and assign individual roles and is represented as a resource stored in the *Forming Team* pool in the form of two coins. The two coins represent an individual task and a team task. The team is expected to work together and produce a learning product, which will be stored in the *TProduct* pool. The team will work on the team task by applying the previous skill represented by *Proficiency* level, which means converting these resources using the *intra-Collaboration* converter into a team effort and storing this in the *TProduct* pool.

The production rate of the team collaboration starts from 1, as shown in the resource connection of the *intra-Collaboration* converter. The collaboration of the team members is expected to be developed over practice and time. So, the rate of the *intra-Collaboration* converter will be increased by some value decided by the game designer/educator in the next team task by the label modifier state connection (dotted arrow with +) between the *TProduct* pool and the resource connection of the *intra-Collaboration* converter.

A delay (a small circle with an hourglass symbol inside) mechanism for a few time steps (st) such as 3 steps is applied to *intra-Collaboration*, showing that team collaboration requires time and also gives individual members time to work on individual tasks. Individual effort is stored in the *IProduct* pool and is produced by applying the skill from the previous proficiency level to the individual task and is represented as the *intra-Competition* converter. The *intra-Collaboration* rate is expected to increase the production of individuals as team members are expected to learn something from the team collaboration. This increase will appear in the production of the next individual task by the label modifier state connection (dotted arrow with +) between the *IProduct* pool and the resource connection of the intra-Competition converter. Again, game designer/educator can specify the amount of increase.

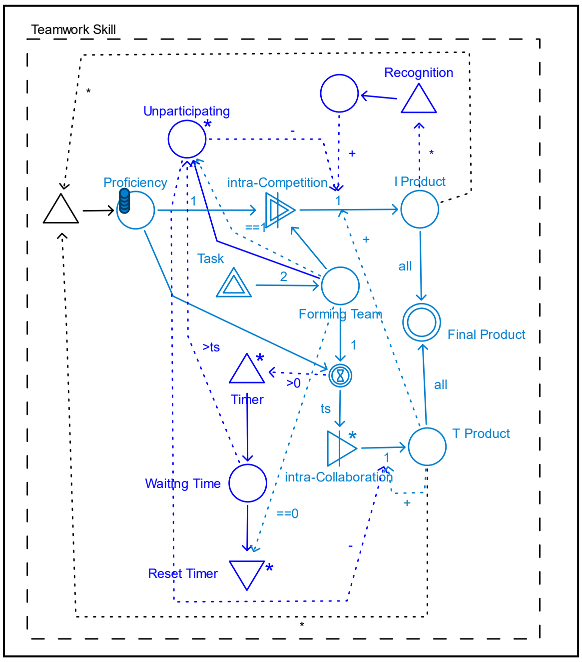


Figure 4.26: Balancing intra- collaboration and competition.

To balance the intra-collaboration and intra-competition, two internal mechanics (recognition and timer) are added, as shown in Fig. 4.26 in dark blue. Each time an individual product is stored in the *IProduct* pool, the individual will receive recognition by the *Recognition* source and this is stored in a pool. The recognition would affect the intra-competition of the team member through the label modifier state connection (dotted arrow with +) between the *Recognition* pool and resource connection of the *intra-Competition* converter. The effect is achieved via increasing the rate of the *intra-Competition* converter by some value as decided by game designers/educators.

To balance this increase of rate by the recognition mechanic, a timer mechanic is applied to count the time (ts) given to individual members such as three time-steps to produce *IProduct* resource. If the condition of the activator state connection (dotted arrow labelled with condition > ts) between the Waiting Time pool and Unparticipating pool is satisfied, this means that the individual member did not complete the individual task. As a result of not participating, the automatic Unparticipating pool will be activated, and the task will be collected from the Forming Team pool. The rate of the intra-Competition converter will be decreased (-) because of the lack of participation of the individual member via the label modifier state connection between the Unparticipating pool and the resource connection of the intra-Competition converter by some value.

In addition, the rate of the *intra-Collaboration* converter will be decreased for the same reason, the lack of participation of the individual member, which will affect team collaboration and production. This is achieved by the same method utilised to decrease the rate of the *intra-Competition* converter through the label modifier state connection between the *Unparticipating* pool and the resource connection of the *intra-Collaboration* converter.

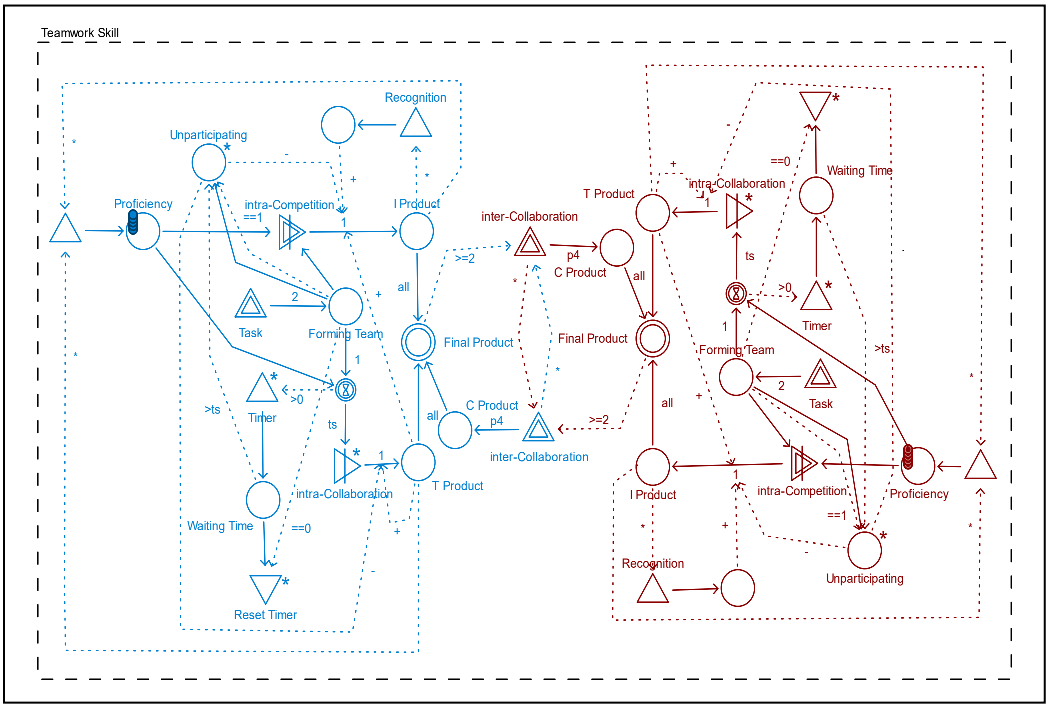


Figure 4.27: Teamwork skill model.

The team and individual production are a sign of improvement of the level of proficiency. The development of the proficiency level is represented by a trigger state connection (dotted arrow labelled with an asterisk * in black) from both the *IProduct* pool and *TProduct* pool to a source that creates the proficiency resource and then this is stored in the *Proficiency* pool to show proficiency improvement.

The interaction between teams should be considered while designing teamwork skill tasks for VFTGs. Fig. 4.27 presents a teamwork skill model between team members and between different teams. The teamwork skill model in Fig. 4.27 shows two teams (blue and red). The second team modelling, which is in red, is identical to the blue team modelling. The inter-collaboration between teams is represented by *inter-Collaboration* sources, one for each team, where a team requests collaboration (clicking on the interactive *inter-Collaboration* source) from the other team. Collaboration between teams depends on the willingness to share with the other team, which is a random factor in the learning process and in the game mechanic. This is represented by *p4* which is a probability chance of willingness to collaborate via the *inter-Collaboration* source.

The collaboration product will be stored in the *C Product* pool. The other team cannot collaborate if it did not develop T Product and I Product. For that the *inter-Collaboration* source should be activated if the *Final Product* pool of the other team contains at least two resources through an activator state connection with (>=2) condition. One team's willingness to collaborate could encourage the other team to collaborate without a request and this is the benefit of the two trigger state connections (dotted arrow labelled with an asterisk) between the two *inter-Collaboration* sources.

The recognition mechanic between teams can be applied in the same way it is applied to one team to encourage and balance collaboration and competition between teams. Inter-collaboration would be rewarded by recognition, which is expected to increase the rate of collaboration and competition. Moreover, unrewarding, which means the other team refused the collaboration request, would decrease the rate. The final step is combining all the efforts, and this is stored in the *Final Product* pool.

4.2.3 World Modelling

World modelling aims to link the field environment to the game environment by setting variables and rules to form the VFTG world. The main variables of VFTG world modelling (displayed in Fig. 4.28) are: space, time, and storytelling. The space and time of the VFTG world construct the spatial and temporal representations of knowledge concepts (Ks), where Ks *are* identified as a part of the task/evaluation model. These spatial and temporal representations of selected Ks enhance the experience of learning by presenting the required knowledge to perform LTs in a different way than the traditional methods of learning where the same *Ks* can be gained from reading books or attending lectures. Therefore, the VFTG world is constructed based on specific time and space associated with the selected Ks.

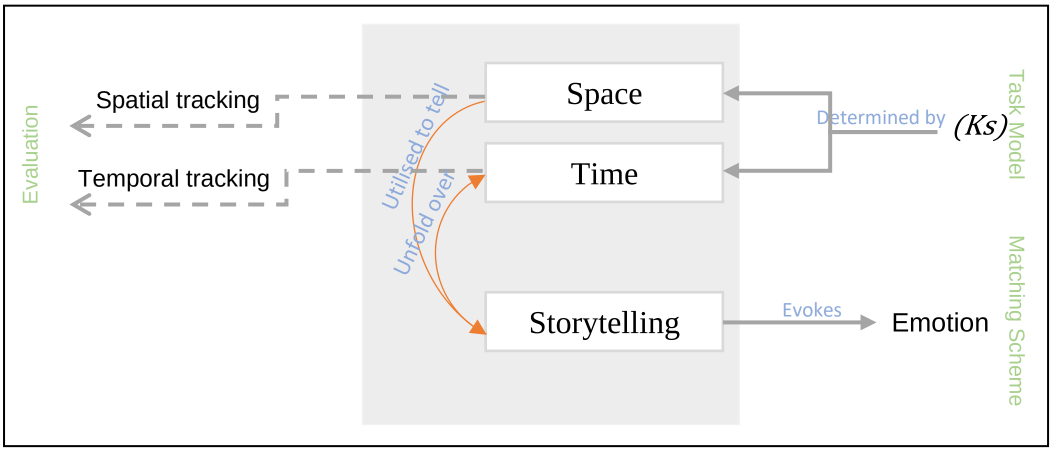


Figure 4.28: World modelling.

The space and time variables of the VFTG world have two different sides: the *Ks* representation side and the well-known side. The well-known side of time consists of seconds, minutes... etc. While the well-known side of space consists of position, coordinates…etc. The player generates a spatial track in the space, which can be used to evaluate the player's abilities as explained in the evaluation model. In parallel, the player creates a temporal track while playing and solving tasks. Temporal tracks should be utilised in the process of evaluation if the time is an important element in performing learning tasks as explained in the evaluation model.

The last variable of VFTG world modelling is storytelling. Types of narrative have been discussed briefly in relation to task complexity; however, storytelling is connected to the time and space of the VFTG world. The story unfolds over the time and utilises spatial objects in the space, as in spatial storytelling. Spatial storytelling depends on organizing some of the objects presented in the VFTG world to indicate the story events to the player; it is like the narrative is spread across the VFTG space. It is common for players to identify their progression by spatial markers (game objects), so VFTG designers should utilise this common culture to spatially communicate Ks to the player inside the VFTG world. Also, Emotion has a significant impact on cognitive processes so storytelling should be utilised to evoke suitable emotion in the VFTG in order to increase motivation and the feeling of presence.

The narrative that would be told during the VFTG is constructed from a number of events where the storytelling begins with a *Start* event and closes with an *End* event and several *Middle* events. The story events can be arranged in different ways, such as linear, linear-branching, or network configurations. Focusing on the spatial storytelling, it can be represented as two different views derived from the first Peuquet's Dual Model (Peuquet, 1988): spatial view and narrative view. Each view consists of an entity, attributes, and relationships. The entity of the spatial view is an object, while the entity of the narrative view is a story event. The attributes of the spatial view are type (fixed or mutable), event, associated knowledge, position, size, etc., while the attributes of the narrative view are type (start, end, middle), embedded emotion if any, and story information. The relationships between entities can be of any kind such as "*part of*" in the case of object entity or "*is following*" in the case of story event entity. The mapping between the two views defines a connection between objects and story events of a scene/level, where a specific object (position) is marked to reveal a story event to players as follows:

$$What \rightarrow Where$$

Given a story event (the "what"), the position (the "where") of the object attached to the event can be identified.

$$Where \rightarrow What$$

Given a position of an object (the "where"), the attached story event (the "what") can be identified.

This storytelling process plays a role in distributing the Ks (the Ks that are employed/gained while performing action on an object) across the game space, which helps the player by:

- Improving their comprehension ability by employing the underlying causal configuration of spatial storytelling to define implicitly relationships between Ks/skills and spatial/event elements.
- Enhancing recalling ability where context and cues support human memory and spatial storytelling draws a comprehensive contextual landscape; then the player can easily recall related Ks/skills when they are located in a meaningful context such as how and why.

World modelling facilitates the three variables to develop the VFTG world in relation to learning and performing learning tasks via Ks (task model), to connect the VFTG world to the evaluation process via spatial and temporal tracking (evaluation model), and to hook the VFTG world to the ELT cycle, especially the first stage – CE – via an emotion element (matching scheme).

The unit of space is an object (O), and the space of VFTG is defined by all the objects that combined to construct the whole VFTG world. These objects can be divided among different scenes/levels. Space (SP) is described as follows:

$$SP = \{O_1, \ldots, O_i, \ldots, O_{|SP|}\} \, for \; 1 \leq i \leq |SP| \qquad (4.19)$$

Where O_i is one object utilised in the process of constructing the space, while $|SP|$ is the number of all objects in the SP.

The unit of time is a game time (gt) which reflects a specific time period. The whole VFTG world can be built based on one gt or several different game times; one for designing each scene/level. For example, gt of the first level could represent the current time or the medieval period. Time (TI) is described as follows:

$$TI = \{gt_1, \ldots, gt_i, \ldots, gt_{|TI|}\} \, for \; 1 \leq i \leq |TI| \qquad (4.20)$$

Where gt_i is one time period utilised to construct one scene/level, while $|TI|$ is the number of all game times utilised to construct the VFTG world.

Scene/level (S) is defined as a set of Os, which is a subset of SP, represented based on a specific gt in addition to the learning tasks presented to the player in that S. Objects need to be also defined according to their main features/attributes such as the spatial view's attributes. Then O is defined by its attributes in a form of a vector:

$$O = [type, K, SE \; position, size, color] \qquad (4.21)$$

Where O is an object, $type$ can be defined as a binary variable (fixed = 0, mutable = 1), SE can be seen as another vector or an index to a story event entity (SE), and K is the associated knowledge concept. Attributes could be added or ignored as needed. For instance, in case of an object does not belong to the spatial view, the SE attribute should be ignored.

Story event entity (SE) is also defined by its attributes as follows:

$$SE = [type, EM, info] \quad (4.22)$$

Where $type$ can be addressed as an element of the set $\{Start, Middle, End,\}$, EM identifies the emotion that SE aims to evoke, and $info$ contains a piece of story information.

Usually, a level represents a section of the game world where it progresses with the difficulty of the tasks, while a scene represents a smaller part, or a slice of the level focused around one idea or one task. However, the design of a scene and level will be treated the same and addressed as:

$$S = gt + \sum_{i=1}^{n} O_i + \sum_{i=1}^{r} LT_i \ , \ gt \in TI \quad (4.23)$$

Where S is a scene or level, gt outlines one element of a TI set that impacts the design of the scene's objects, sum O_i is a subset of SP, and n equals the total number of Os available in the S. Also, $\Sigma\, LT_i$ represent the sum of all learning tasks introduced to the player in that specific S and r equals the total number of LTs.

4.3 Summary

- Contributions:
  1. Assisting the process of designing VFTGs for game designers/educators by modelling the main aspects of FBL and connecting them to game design aspects via three theoretical and conceptual links:
     - ELT modelling to map ELT into the internal economy mechanic.
     - Game modelling links the learning process to gameplay.
     - World modelling connects the field environment to the game environment.
  2. The proposed conceptual framework is built on several learning aspects: ELT, FBL structure (pre, during, post), TBL, task variables (time & location), feedback provision, assessment, some aspects of social learning theories, and high-order skills.
  3. Facilitating the design of in-game assessment and feedback provision in relation to FBL via evaluation model.
  4. Identifying common indicators of players' performance in VFTGs to be standardized.
  5. The models (ELT, evaluation, and skill) are broken down into building blocks to help game designers/educators developing the VFTG design block by block.

- ELT concepts are quantified by the internal economy, which is a core mechanic of designing games, to connect them to game aspects. The three components of internal economy (resources, internal mechanics, feedback loop) are specified for each ELT stage. Building blocks of quantified stages are presented via game machinations showing the learner's progress through the ELT cycle as internal economy. As a result of completing the four stages, four resources are expected to be achieved as the player's gained knowledge/skill (***level1***: observing, ***level2***: reflecting, ***level3***: synthesizing, and ***level4***: doing.)
- Game modelling connects the learning process gameplay through four steps: matching scheme, task model, evaluation model, and world modelling.
- The matching scheme tool helps game designers/educators to adapt the FBL structure to the virtual field in game environment and link it to the required game elements. The tool includes three phases formed based on the FBL structure with adjustments to fit the VFTG environment: pre-, during, and post. Each phase is divided into a number of steps and each step is linked to several game elements to help develop the FT experience in VFTGs. Employing the matching scheme will further develop and link FBL to the game design process.
- TBL is a learning theory that uses the task as a unit of analysis where the syllabus is divided into units (tasks) to arrange learning and the focus is on the process rather than on the final product. The conceptual framework adopts two concepts of TBL by arranging the design of VFTGs around tasks and models the design of tasks to capture the process along with the final product. This leads to players solving a task by analysing it to a series of interactions based on the required knowledge and skills.
- Designing tasks to be solved by analysing them into a sequence of interactions is explained and modelled in the last step of modelling task (task evaluation step-evaluation model).
- Modelling tasks integrates tasks into the VFTG structure and connects them to the game elements that were selected in the previous step (matching scheme). The task model helps to design the pre-task in the *pre-phase* as well as the subtasks in the *during* phase of the matching scheme. Task model consists of three steps: task preparing, task design, and task evaluation.
  - In Task preparing, game designers/educators define some learning elements (goal, objectives, complexity, and difficulty) that are required for designing the learning tasks of VFTGs.

- In task design step, game designers/educators identifie the game mechanics and learning outcomes and then link them to selected game elements based on the matching scheme. Examples of game mechanics that can be employed to design learning tasks based on two types in relation to ELT: grasping experience (CE & AC) and transforming experience (RO & AE) are provided.
- The last step is task evaluation where assessment is built upon clear learning elements and blended with the components of the task design step. The task evaluation process is explained via the evaluation model where the process is modelled as internal economy mechanic which produces an essential resource which is the indicator.

- This proposed conceptual framework sees assessment as an integrated part of VFTGs to ensure the effectiveness of feedback. The advanced technology of VFTGs overcomes the limitations of physical FTs where assessment takes place days or weeks after completing the FT.
- The evaluation model can be utilised as a building block which can be attached after each stage of ELT or used once at the end of the cycle. Indicators reference evidence of gaining knowledge/skills based on player's interactions within the VFTG. Learning tasks can be very unique to a particular game or subject. Yet, defining learning tasks according to the expected series of interaction to solve leads to the standardisation of assessment based on captured indicators. Three types of indicators are distinguished for evaluation in VFTGs: knowledge indicator, FB indicator, and skill indicator. The series of player's interactions that lead to completing the learning task is defined and utilised to identify these three types of indicators.
- A skill task is a composite of learning tasks as steps defined based on a specific theory or model. Modelling skills as internal economy mechanic supports the connections between learning skills and game design. Two important skills (teamwork & decision-making) for the 21st century workplace/field are modelled, and game designers/educators can adapt to design more skills.
- The teamwork skill model considers some aspects of social learning theories such as constructing knowledge with other players (learning of practice) and observing other players' experiences (learning of meaning).
- World modelling aims to link the field environment to the game environment by setting variables and rules to form the VFTG world. The main variables of VFTG world modelling are: space, time, and storytelling. The knowledge concept helps

to determine the time and space which are necessary for the spatial and temporal representations to enhance the learning experience in an authentic environment. While the storytelling unfolds over time and utilises spatial objects in the space. The essential aim of storytelling is to evoke emotions, which are supposed to be specified by the matching scheme tool, to encourage motivation and presence, in addition to placing Ks and skills in a meaningful context to support the player's recalling of the learning experience. The player's performance creates spatial and temporal tracks while playing. These tracks should be utilised to evaluate the player's abilities as needed via the evaluation model.

- The selected learning aspects are employed for different FTs of different subjects in the literature such as social studies (Djonko-Moore & Joseph, 2016), marketing education (Frontczak, 1998), legal education (Higgins, Dewhurst, & Watkins, 2012), geology (Orion & Hofstein, 1994), environmental education (Jose, Patrick, & Moseley, 2017), tourism education (Sofield & Marafa, 2019), and geography (Dummer et al., 2008).
- The conclusion from the last previous two points, the proposed conceptual framework can be generalised to design and develop VFTGs of any subjects that can utilise the learning pedagogies that this framework is built on.

Chapter 5

5 Preliminary Study: Data Analysis and Results

5.1 Introduction

Following the construction of the first version of the conceptual framework, a preliminary study was conducted as the first method of evaluation and to approve the effectiveness of the proposed framework. The first version included three links that connect FBL to game design (game elements and mechanics): the ELT modelling (Section 4.2.1), matching scheme (Section 4.2.2.1), and initial perception of world modelling (Section 4.2.3). A questionnaire was used to collect the data, which was created specifically for this study. The research ethical approval was obtained from the Ethics Representative of the Department of Computer Science before sending invitations via email. Also, the questionnaire was collected anonymously without any private or identity information. The participants were informed that their answers to the questionnaire will be used for research purposes, will be removed when they are no longer needed for this research, and the collected data will not be shared with any third party. The main purpose of the questionnaire was for experts to evaluate the conceptual framework and provide feedback for its improvement.

5.2 Method

The preliminary evaluation study of the first version of the conceptual framework was designed to explore the experts' perceptions of its connection, usefulness, and usability. The goal was to obtain experts' feedback in order to identify issues before starting the prototyping and quasi-experiment. A questionnaire and summary of the first version of the framework were administered via Bristol Online Survey. Invitations were sent via email and followed up with reminders a week later. Game designers/educators from universities and schools were the target participants. The questionnaire aimed to address the following questions:

Is the framework easy to apply by both game designers/educators (usability)?
Can the framework provide a connection between learning aspects and game design aspects?

5.3 Questionnaire

After identifying what the questionnaire was expected to measure, a draft questionnaire was designed. The second step was for three experts to assess the questionnaire: two professors from the University of Durham and one from Umm Al-Qura University. The three professors identified themselves as males with more than 10 years of experience. Also, all of them work at universities, and two of them defined themselves as game designers/ educators and one as an educator. Based on the experts' feedback, some changes were made to the wording of the questions. A piloting phase was then followed with five representative participants. Four of the representative participants are from Saudi Arabia while one participant is from the UK. One participant works in a school and the rest work at universities. Three of the participants defined themselves as educators/game designers, one participant as a game designer and one as an educator. Regarding to gender, one participant identified herself as a female and the rest as male. With respect to years of experience, one participant stated that he/she had less than five years of experience, two participants stated they had five to ten years of experience, and two participants reported that they had more than ten years of experience. A major change in the structure of the questionnaire was made in addition to a few changes to the wording of the questions. The questionnaire included an introductory message along with researcher information and privacy and data protection details. A detailed explanation of the first version of the conceptual framework was provided to the participants in the form of PowerPoint slides in addition to the three important aspects to facilitate understanding of the conceptual framework:

- Internal economy: game mechanics.
- Kolb's Experiential Learning Theory (ELT).
- Game machinations.

The participants were instructed to go through the slides before answering the questionnaire, which would it takes twenty to thirty minutes. As mentioned before, the invitation emails were sent to participants, and after one week followed by reminder emails. All the responses were collected during four weeks after sending the reminder emails. The questionnaire gathered data via six sections (see Appendix A) to answer the preliminary study questions. The first section included questions to obtain demographic information on sex and location while the second section contained questions about the professional experience (occupation, years of experience, type of learning institution, and duties). The third section consisted of questions about expertise regarding digital games (DGs), field-based learning (FBL), and virtual field trip games (VFTG). The connection and usefulness evaluations section formed the fourth section. This section included questions about each link of the framework to help measure the usefulness of the framework and the connection between learning and game design aspects. The fifth section contained questions to measure the usability of the framework, which was a modified version of the System Usability Scale (SUS). The sixth and last sections included questions about the overall framework and suggestions.

The demographic information section contained two questions, while the professional experience included four questions. Questions in the expertise and fourth sections required answers based on a five-point Likert scale (strongly Agree=5, Agree=4, Neutral=3, Disagree=2, Strongly Disagree=1). The expertise section had nine statements in total and the fourth section consisted of fifteen questions. It is explained in the questionnaire, the framework includes three links, each of which provides a different connection between learning aspects and game design and is measured based on ten statements. While the usefulness of each link was measured based on five statements. The SUS (Brooke, 1996) was used, which is a well-established questionnaire for measuring usability. In this study, the fifth section contained the SUS ten items with a Likert attitude scale to enable subjective evaluation of usability. The SUS was utilised in this study even though the framework is not in the form of a digital artefact because the participants were instructed in clear language not to start answering the questionnaire unless they were willing to spend time sketching the design of a VFTG following the framework and they can ask the researcher any questions via emails. Especially since the first version was simple and involved a few steps and designing concepts, the interaction with the framework would require reasonable time from the participants. Literature (Arnab et al., 2015; Carvalho et al., 2015) shows the use of SUS to measure the usability of game designing frameworks in a similar setting. Where participants were asked to use the proposed model/framework to design GBL or SG either in their own time or in an organised setting as a workshop and this setting could be considered for future work. The final section included four questions, two of them are open-ended questions. The questionnaire was administered using Bristol Online Survey and the link was sent via email along with invitations. The following sections explain the statistical tests used to analyse the collected data, results, and findings, respectively.

5.4 Statistical Analysis

Exploratory data analysis (EDA) was conducted, using appropriate statistical measures for each variable to calculate descriptive statistics such as frequencies and percentages. Reliability analysis was applied (Cronbach's alpha) to indicate the consistency and reliability of the statements for each scale. Likert-scale questions were analysed by calculating the weighted average for each statement to obtain data on the level of agreement. Hypothesis testing was employed to inspect the differences across the following variables: occupation, years of experience, type of learning institution, and gender, in terms of connection, usefulness, and usability. The Mann-Whitney U Test was applied to analyse the data on learning institution and gender, while Kruskal-Wallis and Dunn Post hoc tests were employed on the other variables. The significance level was set to 0.05 for all comparisons, and all statistical analyses were performed in IBM SPSS Statistics (version 26 for Windows, IBM Corporation).

5.5 Results

A total number of 23 participants answered the given questionnaire. There were only two missing values regarding the years of experience of the participants. Discarding any case with missing values may affect the representativeness of the results, so missing values were imputed by mode (the most commonly observed value in the variables) because the variables were categorical. Table 5.1 displays a summary of demographic information and information on experience. The demographic information shows that there were 7 females (30.4%) and 16 males (69.6%) in the sample while all the participants were from the UK.

The information on experience summarises the following: occupation, years of experience, type of learning institution, and duties. Overall, there were ten educators (43.5%), seven game designers (30.4%) and six participants who defined themselves as both game designers and educators (23.1%). With respect to years of experience, six participants (26.1%) stated that they had less than five years of experience, five participants (21.7%) stated they had five to ten years of experience, and ten participants (43.5%) reported that they had more than ten years of experience. Regarding learning institution, five participants work in schools (21.1%) and 18 work at universities (78.3%). In terms of their duties, eight were involved in developing learning applications (34.8%), four had administrative/management duties (17.4%), seven had teaching roles (30.4%), and four were researchers (17.4%).

Table 5.1: Characteristics of preliminary study variables.

Variable	*Frequency*	*Percentage*
Gender		
Male	16	69.57%
Female	7	30.43%
Years of experience		
Less than 5 years	6	26.09%
5–10 years	5	21.74%
More than 10 years	12	52.17%
Occupation		
Educator	10	43.48%
Game designer	7	30.43%
Both	6	26.09%
Learning institution		
School	5	21.74%
University	18	78.26%
Duties		
Involved in developing learning applications	8	34.78%
Administrative/management duties	4	17.39%
Teaching role	7	30.43%
Researcher	4	17.39%

Cronbach's alpha was calculated for the four scales and the results are summarised in Table 5.2: connection (α= 0.805), usefulness (α= 0.718), and usability (α= 0.869). The results indicate good internal consistency and reliability for the statements of each scale.

Table 5.2: Cronbach's alpha reliability analysis.

Scale	***Number of items***	***α***
Connection	10	0.805
Usefulness	5	0.718
Usability	10	0.869
Suggestions	3	0.834

The statements on expertise revealed insights about participants' knowledge regarding digital games, FBL, and VFTG. Table 5.3 shows the descriptive statistics on expertise. Regarding DGs experience, most participants indicated agreement/strong agreement with having "played digital games" (N=20, 86.96%), which shows that most of the participants had experience with playing digital games. More than half of the participants agreed or strongly agreed regarding "designing digital games" (N=13, 56.52%), which is to be expected given that the number of game designers who participated exceeded the number of educators. However, the agreement percentage dropped under 50% with regard to designing educational DGs (N=10, 43.48%) and the majority of participants (N=13, 56.52%) expressed strong disagreement or disagreement. In contrast, a higher percentage of participants agreed or strongly agreed that they used educational games in teaching (N=16, 59.56%) while only seven participants (N=7, 30.43%) expressed strong disagreement or disagreement. Most participants seemed to agree that they had knowledge of DGs (Median= 3.75, IQR= 4.38–2.88), implying that their knowledge of DGs is higher than that of FBL, and VFTG as will be shown below.

Table 5.3: Descriptive statistics on expertise with DGs, FBL, and VFTGs.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Median	IQR
I have played digital games.	1 4.35%	1 4.35%	1 4.35%	6 26.09%	14 60.87%	5.00	5–4
I have experience with designing digital games.	6 26.09%	4 17.39%	0 0.00%	1 4.35%	12 52.17%	5.00	5–1
I have used educational games in teaching.	3 13.04%	4 17.39%	0 0.00%	8 34.78%	8 34.78%	4.00	5–2
I have experience with designing educational digital games.	7 30.43%	6 26.09%	0 0.00%	6 26.09%	4 17.39%	2.00	4–1
Knowledge of DGs						3.75	4.38–2.88
I am familiar with field-based learning	2 8.70%	7 30.43%	0 0.00%	7 30.43%	7 30.43%	4.00	5–2
I am familiar with Experiential Learning Theory.	3 13.04%	6 26.09%	1 4.35%	6 26.09%	7 30.43%	4.00	5–2
I have experience with applying field-based learning.	4 17.39%	6 26.09%	0 0.00%	11 47.83%	2 8.70%	4.00	4–2
Knowledge of FBL						3.67	4.17–2.17
I am familiar with virtual field-trip games.	7 30.43%	4 17.39%	0 0.00%	7 30.43%	5 21.74%	4.00	4–1
I have experience with designing virtual field trip games.	10 43.48%	6 26.09%	1 4.35%	4 17.39%	2 8.70%	2.00	4–1
Knowledge of VFTGs						2.5	3.75–1

Regarding knowledge of FBL, the majority of participants stated agreement or strong agreement with having familiarity with FBL (N=14, 60.86%), and a similar percentage expressed agreement or strong agreement regarding “experience with applying field-based learning” (N=13, 56.53%). Again, the majority (N=13, 56.52%) specified that they agreed or strongly agreed with having familiarity with ELT. Overall, participants agreed that they had knowledge of FBL (Median= 3.67, IQR= 4.17–2.17). Participants’ knowledge of FBL was higher than their knowledge of VFTGs.

The final index of knowledge was regarding VFTGs. Participants seemed to be divided regarding familiarity with VFTGs as some participants expressed strong disagreement or disagreement (N=11, 47.82%) while others indicated that they agreed or strongly agreed (N=12, 52.17%). On the other hand, most participants (N=16, 69.57%) indicated strong disagreement or disagreement regarding having experience in designing VFTGs. The overall knowledge of VFTGs indicates a disagreement trend (Median= 2.5, IQR= 3.75–1).

5.5.1 Connection

The descriptive statistics for the connection scale are displayed in Table 5.4. All statements have the same median (4) and the IQR values range from 0 to 2, which indicates that most participants expressed agreement regarding each link of the framework. That leads to an overall connection score that indicates an agreement trend (Median= 4.20, IQR= 4.57–3.79).

A Kruskal-Wallis (KW) test was conducted to test whether there was a difference in connection score regarding variables with more than two groups: occupation, years of experience, and duties. The null hypothesis states there is no significant difference in connection score regarding one of these variables while the alternative hypothesis states there is a significant difference. According to the obtained results in Table 5.5, there is no evidence to accept the null hypothesis with 95% confidence. This indicates that significant differences (H(2) = 15.161, P = 0.001) were found among the three levels of occupation (educator, game designer, and both) with a mean rank of 17.25 for both, 16.43 for game designer, and 5.75 for educator.

Table 5.4: Descriptive statistics on connection.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Median	IQR
Kolb's Experiential Learning Theory enhances the design of virtual field trip games.	0	1	2	17	3	4.00	4–4
	0.00%	4.35%	8.70%	73.91%	13.04%		
The first link connects learning concepts to game mechanics (internal economy).	0	3	4	6	10	4.00	5–3
	0.00%	13.04%	17.39%	26.09%	43.48%		
Graphical representations (modelling ELT stages as internal economy) can accurately transpose concepts into practice.	2	1	4	9	7	4.00	5–3
	8.70%	4.35%	17.39%	39.13%	30.43%		
Field-based learning structure of three phases (pre-, during, post-) enhances the design of virtual field trip games.	0	0	4	13	6	4.00	4.5– 4
	0.00%	0.00%	17.39%	56.52%	26.09%		
The second link connects learning concepts to game elements.	0	2	0	11	10	4.00	5–4
	0.00%	8.70%	0.00%	47.83%	43.48%		
Graphical representation (the table) can accurately transpose concepts into practice.	0	3	0	11	9	4.00	5–4
	0.00%	13.04%	0.00%	47.83%	39.13%		
The environment settings (realism, multi-role, multimodal of interaction, and complexity)	0	0	2	13	8	4.00	5–4
	0.00%	0.00%	8.70%	56.52%	34.78%		

enhances the design of virtual field trip games.							
The third link connects learning concepts to game mechanics (internal economy).	0	1	4	10	8	4.00	5–4
	0.00%	4.35%	17.39%	43.48%	34.78%		
The third link connects learning concepts to game elements such as immersion, identity selection, points, and narrative.	0	0	1	14	8	4.00	5–4
	0.00%	0.00%	4.35%	60.87%	34.78%		
The third link can transpose concepts into practice.	0	0	3	13	7	4.00	5–4
	0.00%	0.00%	13.04%	56.52%	30.43%		
Connection score						4.20	4.55–3.65

In order to determine which occupations were significantly different from each other in terms of the connection scores, a multiple comparisons procedure was conducted, controlling for Type I error across tests by using the Bonferroni approach. The results of the Mann-Whitney test indicated that the connection score was greater for game designers (mean rank = 14) than for educators (mean rank = 5.50), U= 0.0, p < 0.001. Also, the results of the Mann-Whitney test indicated that the connection score was greater for both (mean rank = 13.08) than for educators (mean rank = 5.75), U = 2.5, p = 0.001.

Table 5.5: KW test – connection and occupation.

	Connection
Kruskal-Wallis H	15.161
df	2
Asymp. Sig.	.001

Regarding years of experience, there is no evidence to reject the null hypothesis with 95% confidence, which indicates no significant differences (KW = 5.978, P= 0.050, df= 2) were found among the three levels of years of experience. The mean rank for these levels were as follows: 17.25 for “less than 5 years”, 11.25 for “more than 10 years”, and 7.50 for “5–10 years”.

A Mann-Whitney U Test was conducted to test whether there was a difference in connection score regarding learning institution. The null hypothesis is that the distribution of scores for the two groups of learning institution are equal, while the alternative hypothesis is that the mean ranks of the two groups are not equal. According to the obtained results in Table 5.6, there is no evidence to accept the null hypothesis with 95% confidence, which indicates a significant difference (U = 17.50, P = 0.039) was found among the two levels of learning institution (school and university) with mean ranks of 13.53 for university and 6.50 for school.

Table 5.6: Connection and learning institution.

	Connection
Mann-Whitney U	17.500
Wilcoxon W	32.500
Z	-2.056
Asymp. Sig. (2-tailed)	.040
Exact Sig. [2*(1-tailed Sig.)]	.037[b]
Exact Sig. (2-tailed)	.039
Exact Sig. (1-tailed)	.020

To test the difference in connection scores across gender, a Mann-Whitney U Test was conducted. The null hypothesis is that the distribution of scores for the two groups of gender are equal, while the alternative hypothesis is that the mean ranks of the two groups are not equal. The result reveals that there is no evidence to reject the null hypothesis with 95% confidence, which indicates there is no significant difference (U = 48.50, P = 0.62) between males and females regarding connection score with a mean rank of 12.47 for males and 10.93 for females.

5.5.2 Usefulness

Table 5.7 displays the descriptive statistics for the usefulness scale. The results show that the median for all statements ranges between 4 and 5 and the IQR varies between 1 and 2, which reveals that most participants indicated agreement. It is worth noting that all participants (N=23, 100%) indicated agreement or strong agreement with the second link being understandable (Median = 5, IQR= 5–4), and also most participants (N=22, 95.66%) indicated agreement or strong agreement with the third link being understandable (Median = 5, IQR = 5–4). The overall median of the usefulness scale (Median = 4.60, IQR= 4.70–3.80) indicates that it has a strong agreement trend.

Table 5.7: Descriptive statistics on usefulness.

Statement	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Median	IQR
The first link is understandable.	2	4	0	8	9	4.00	5–3
	8.70%	17.39%	0.00%	34.78%	39.13%		
Graphical representations (modelling ELT stages as internal economy) are explicit	1	2	4	8	8	4.00	5–3
	4.35%	8.70%	17.39%	34.78%	34.78%		
The second link is understandable.	0	0	0	10	13	5.00	5–4
	0.00%	0.00%	0.00%	43.48%	56.52%		
Graphical representation (the table) is explicit.	0	0	0	13	10	4.00	5–4
	0.00%	0.00%	0.00%	56.52%	43.48%		

The third link is understandable.	1	0	0	6	16	5.00	5–4
	4.35%	0.00%	0.00%	26.09%	69.57%		
Usefulness score						4.60	4.70–3.80

The first hypothesis testing performed was to test whether there was a difference in usefulness score across the different levels of occupation by conducting a Kruskal-Wallis test with hypotheses similar to those used in the connection and occupation tests. According to the obtained results in Table 5.8, there is no evidence to accept the null hypothesis with 95% confidence, which indicates significant differences (KW(2) = 16.927, P < 0.001) were found among the three levels of occupation (educator, game designer, and both) with a mean rank of 17.93 for game designer, 15.92 for both, and 5.50 for educator.

Table 5.8: K-W test – usefulness and occupation.

	Usefulness
Kruskal-Wallis	16.927
df	2
Asymp. Sig.	.000

In order to determine which occupations were significantly different from each other in terms of usefulness scores, a multiple comparisons procedure was conducted, controlling for Type I error across tests by using the Bonferroni approach. The result of the Mann-Whitney test indicated that the usefulness score was greater for game designer (mean rank = 14) than for educator (mean rank = 5.50), U= 0.0, p < 0.001. Also, the result of the Mann-Whitney test indicated that the usefulness score was greater for both (mean rank = 13.50) than for educator (mean rank = 5.50), U= 0.0, p < 0.001.

On the other hand, no significant differences (KW(2) = 5.786, P = 0.055) were found among the three levels of years of experience (less than 5 years, 5–10 years, and more than 10 years) with 95% confidence. The mean ranks were 16.25 for “less than 5 years”, 12.17 for “more than 10 years”, and 6.50 for “5–10 years”.

The two levels of learning institution were tested to determine whether there was a difference in relation to usefulness scores by conducting a Mann-Whitney U Test with similar hypotheses to those previously tested in Section 5.5.1. The results in Table 5.9 show that there is no evidence to accept the null hypothesis with 95% confidence, which indicates a significant difference (U = 7, P = 0.002) was found among the two levels of learning institution (school and university). The mean ranks were 14.11 for university and 4.40 for school.

Table 5.9: Usefulness and learning institution.

	Usefulness
Mann-Whitney U	7.000
Wilcoxon W	22.000
Z	-2.866
Asymp. Sig. (2-tailed)	.004
Exact Sig. [2*(1-tailed Sig.)]	.002[b]
Exact Sig. (2-tailed)	.002
Exact Sig. (1-tailed)	.001

Regarding the last variable to be analysed, gender, there is no significant difference (Mann-Whitney U= 39.50, P= 0.278) between males, and females in usefulness scores. Also, the mean and sum of ranks for males (13.03) are higher than females (9.64).

5.5.3 Usability

The usability score was calculated based on the System Usability Scale (SUS). The SUS score is calculated by totalling the scores of each statement from the SUS scale. Each statement's score will vary from 0 to 4, and for the odd-numbered statements (1, 3, 5, 7 and 9) the calculated score is the scale position minus 1. For the even-numbered statements (2, 4, 6, 8 and 10), the calculated score is 5 minus the scale position. The next step involves totalling the calculated scores of all ten statements and multiplying the total by 2.5 in order to create an overall SUS value which ranges from 0 to 100.

The average usability SUS score with μ= 68.043, which indicates that the usability performance is better than average. The occupation and learning institution variables show similar effects on the SUS score as on connection and usefulness. The result for the occupation variable is ($KW(2) = 15.736$, $P < 0.001$) for the three levels of occupation (educator, game designer, and both). To verify which occupations were significantly different from each other in terms of usability, a multiple comparisons procedure was conducted, controlling for Type I error across tests by using the Bonferroni approach. The result of the Mann-Whitney test indicated that the SUS score was greater for game designers (mean rank = 13.64) than for educators (mean rank = 5.75), U= 2.5, $p < 0.001$. Also, the result of the Mann-Whitney test indicated that the connection score was greater for both (mean rank = 13.33) than for educators (mean rank = 5.60), U= 1.0, $p < 0.001$. The result for the learning institution variable is U = 5.0, P = 0.001 for the two levels of learning institution (school and university).

The result for years of experience (Kruskal-Wallis(2) = 6.668, P= 0.036) for the three levels of years of experience variable (less than 5 years, 5–10 years, and more than 10 years) showed signs of significant difference. However, the multiple comparisons procedure conducted using the Mann-Whitney U Test, controlling for Type I error across tests by using the Bonferroni approach, showed no significant difference for all three levels of years of experience. The last result for the gender variable is Mann-Whitney U= 37.0, P= 0.216 for the two levels of gender (males and females).

5.5.4 Suggestions

The last section of the questionnaire included three statements to evaluate the overall conceptual framework. Table 5.10 shows the statements and summarises the descriptive analysis of the overall evaluation. A high percentage of participants (N=20, 86.96%) stated agreement or strong agreement regarding the conceptual framework being comprehensive of the required FBL concepts and believing the conceptual framework will help connect FBL to game designing. This percentage decreased (N=16, 69.56%) regarding the applicability of the graphical representations. However, the overall evaluation score (median = 4.33, IQR = 4.67–4) indicates strong agreement with the overall conceptual framework.

Table 5.10: Descriptive statistics for participants' suggestions

Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree	Median	IQR
I consider the conceptual framework to be comprehensive of required FBL concepts.	1	1	1	10	10	4.00	5–4
	4.35%	4.35%	4.35%	43.48%	43.48%		
I believe the conceptual framework will help connecting FBL to game designing.	0	1	2	9	11	4.00	5–4
	0.00%	4.35%	8.70%	39.13%	47.83%		
I consider the graphical representations of the framework to be applicable.	0	5	2	8	8	4.00	5–3
	0.00%	21.73%	8.69%	34.78%	34.78%		
Overall						4.33	4.67–4

The participants answered a question (Fig 5.1) regarding the possibility of considering using the framework in their future research. The highest percentage (39.13%) of participants indicated that they would definitely use the framework. This was followed by 30.43% who would probably use it, and 26.09% who would probably not use it. Only

4.35% indicated that they would definitely not use it. The common reason stated by participants in an open-ended question who indicated they would not use the framework is that they do not do research in this area, while one participant expressed the intention to use the framework to evaluate off-shelf VFTGs.

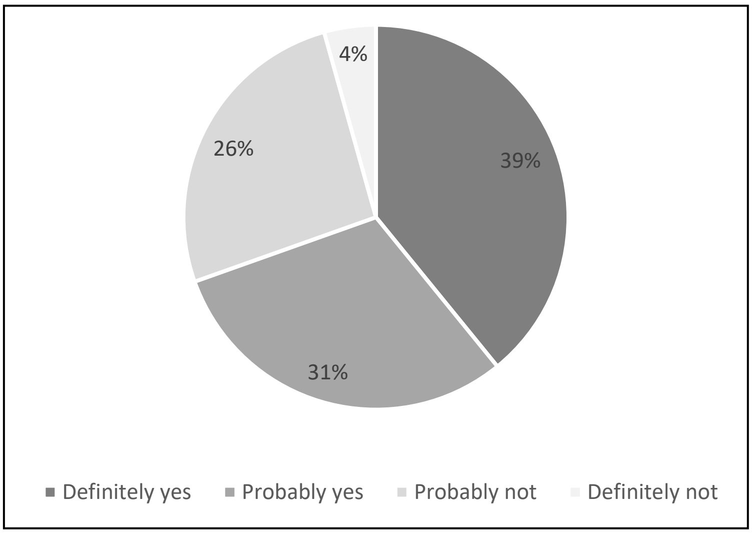


Figure 5.1: Using the framework in future research.

Participants were asked to complete an open-ended question asking whether there was any concept missing from the framework that would support designing VFTG games. Some of them mentioned skills while others said that it was very helpful, but examples were needed to improve understanding and demonstrate how learning activities and outcomes might be applied. In addition, the assessment was mentioned by educators as an important concept to be considered while designing VFTG. However, some of the participants decided not to identify any concepts as they did not have enough knowledge to suggest new concepts.

5.6 Findings

The participants in general found the first version of the framework useful and usable and indicated that it connects learning aspects with game design. However, the strongest opinions were associated with the usefulness scale of the conceptual framework. Participants' opinions were affected by their occupations and learning institutions. The participants who identified themselves as game designers or both (game designer and educator) found the framework to be useful, usable, and provided a connection between FBL and game design aspects to a greater extent than did educators. On the other hand, gender and years of experience did not show any effect on these scales. Based on the overall evaluation and the possibility of using the framework in future research, the proposed framework shows promising credibility with regard to its effectiveness.

The perceptions of participants got affected by only the occupations (game designer/ both and educator) and the differentiation of these two groups' opinions is considered modest. The concepts or symbols of game machinations may play a role in that small differentiation. In general, the lack of knowledge of game design (internal economy) seems to impact the participants' opinions regarding the connections, usefulness, and usability of the first version of the conceptual framework.

Based on the findings of this study, the conceptual framework was improved, and the full version is presented in Chapter 4. The improvements included the following points:

- The second link is extended to link the learning process to the gameplay (game modelling) to cover tasks, evaluations, and skills.
- The Task model is structured to link some learning elements to game design aspects.
- The evaluation model is constructed. The educators shed a light on the importance of assessment in FBL.
- The skill model is added. The higher-order skills are important but only two skills (decision-making skill and teamwork skill) were selected to be modelled as they are required in the field.
- As the participants with a game designer/both occupation and those with an educator occupation differ a little bit in their perceptions of the first version of the framework, the consideration is taken to employ the game machinations along with a well-known method (mathematical model) to define an essential addition to the first version which is defining learning tasks based on interactions and indicators resource.
- The world modelling is improved to follow the same method of defining tasks and indicators, which leads to a greater connection between the learning tasks and the VFTG world.

5.7 Summary

- The preliminary study shows encouraging results regarding the connection, usefulness, and usability of the first version of the conceptual framework.
- The proposed conceptual framework is improved based on the participants' suggestions to include the task model, evaluation model, and skill model.
- The world modelling is improved.

- The results of this preliminary study provide the answer to the third question of the thesis (***To what extent can the proposed conceptual framework build a linkage between FBL aspects and game design aspects?***) as the overall connection score indicates an agreement trend and the overall participants found the proposed framework provides connection and at the same time is useful and usable.

Chapter 6

6 Virtual Field Trip Game Prototyping

6.1 Overview

This chapter explains the process of producing a virtual field trip game (VFTG) prototype to be utilised in the research experiment (Chapter 7). The prototype is a VFTG called Island of Volcanoes, which is set in the island of Bali. The prototyping process has three main steps: inclusion of learning aspects, prototype design, and prototype implementation. The learning aspects, including the learning pedagogies and content utilised in the VFTG prototype, will be explored first. A detailed description will then be presented of the prototype design based on the proposed conceptual framework (Chapter 4). Finally, the key technical steps involved in implementing the VFTG in Unity 3D Engine will be discussed.

6.2 Learning Aspects

Game-based learning (GBL) is provided by the VFTG prototype, meaning that it contains learning pedagogies and content. This section covers the learning aspects included in the prototype design and implementation.

The learning pedagogies follow the proposed conceptual framework (Chapter 4), which was developed to link ELT, the modelling task, higher-order skills, evaluation, and feedback to the game mechanics and elements. Section 6.3 on prototype design provides a detailed explanation of the learning pedagogies applied to the design of the Island of Volcanoes VFTG.

Learning content was selected before the Covid-19 pandemic. Secondary school students seem to benefit more from a VFTG that provides an authentic learning experience through a virtual field trip (VFT) than geography university students. Geography departments worldwide apply different methods of teaching and learning to support students, including lectures, workshops, and especially fieldwork. However, the pandemic limited fieldwork opportunities for students around the world and transformed people’s perspectives on teaching and learning. If university-level learning content had been selected, a VFTG would have been designed to explore and develop possible solutions to contemporary crises (political violence and terrorism; migration and refugees; race and indigenous struggles). A potential VFTG would create an authentic VFT experience of crises, improving awareness of their multiple dimensions and effects, as well as enhancing critical and analytical skills in geographical reasoning and planning/evaluating potential solutions. A VFTG could also be designed to explore natural hazards, risk and resilience, providing an authentic and rich learning environment with real-world-like scenario assessment tasks to understand natural hazards (earthquakes, volcanic, tsunami) and risk assessment. The assessment could focus on the high-order skills (analytical skills, communication skills for a general audience, management skills) involved in implementing conceptual and practical disaster management strategies.

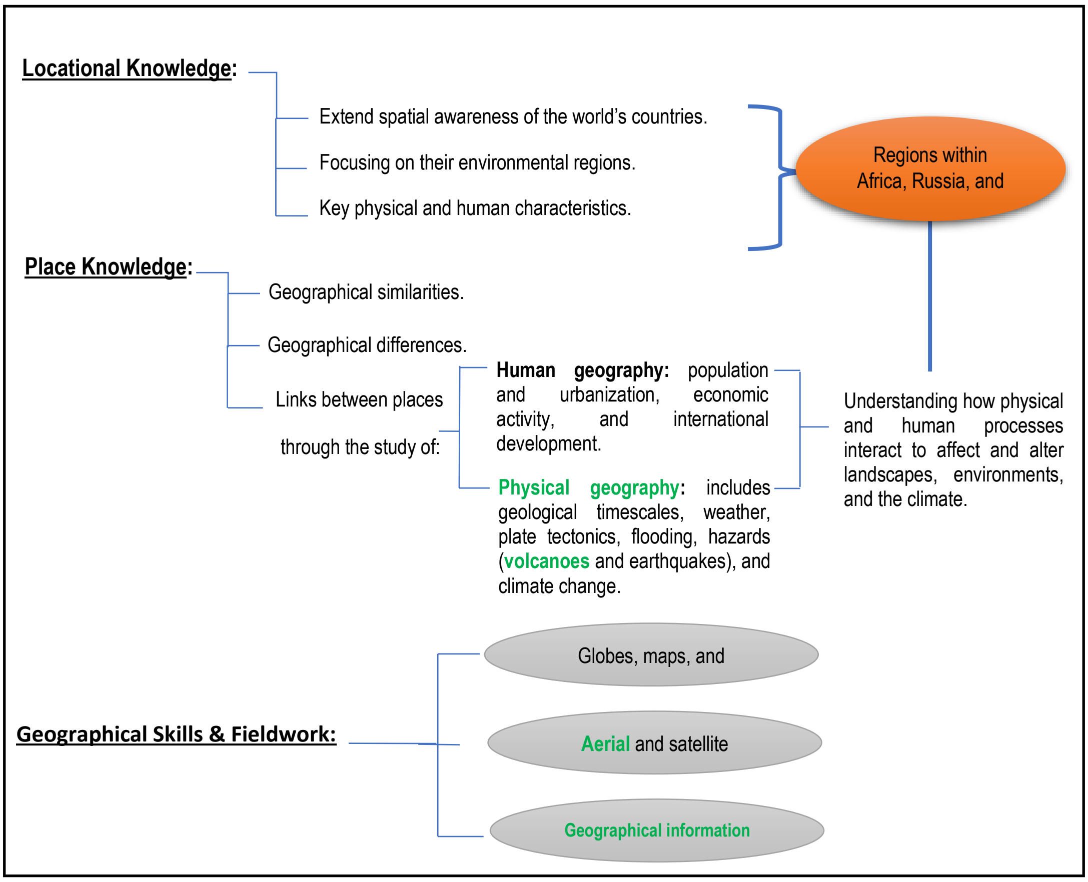


Figure 6.1: KS3 Geography curriculum

However, the VFTG prototype in the current study (Island of Volcanoes) was instead designed and implemented to provide secondary school students with a VFT experience. Learning content was chosen from the most recent National Curriculum in England for Geography Key Stage 3 (KS3). The curriculum builds students' knowledge about locational knowledge, place knowledge, and human and physical geography. In Fig. 6.1, the concepts presented in green text are included in the design of the VFTG.

Decision-making is a high-order skill that was considered when designing and developing the Island of Volcanoes VFTG based on the conceptual framework (Chapter 4). This skill is required in the 21st century, especially in field trips (FTs) and the workplace. Geographical skills were also incorporated into the prototype, such as using aerial imagery and gathering geographical information.

6.3 Prototype Design

The Island of Volcanoes prototype design followed the proposed conceptual framework to connect field-based learning (FBL) to the game design (game mechanics/elements) through creating the following three links:

- Mapping ELT to the internal economy mechanic (ELT modelling).
- Transforming the learning process to gameplay (Game modelling).
- Integrating the field environment to the game environment (World modelling).

6.3.1 The First Link (ELT Modelling)

The VFTG is designed to provide a FT experience of physical geography and natural hazards. First, the ELT cycle is understood based on ELT modelling to formulate the resources of each stage and their flows as an internal economy along with the required internal mechanics and feedback loops. The components of ELT modelling are summarized in Table 6.1, followed by detailed explanations.

Table 6.1: The components of ELT modelling.

Stage	Required Resources	Produced Resource	Internal Mechanic	Feedback Loop
CE	Previous Knowledge	*Level1* – observing volcanoes in the island.	Creating (source)	Positive feedback loop: the more the player puts out fires, the more he/she observes the island.
RO	*Level1*	*Level2* – Collecting information about observed volcanoes.	Creating (source)	Positive feedback loop: if the player finds one piece of information, the rest of the required data can be found easily.
AC	*Level2*	*Level3* - Hypothesizing and planning from observations and collected data. + *Level2*	Converting (converter)	Possible negative feedback loop, where it will get harder to hypothesize and plan with the appearance of new signs of natural hazards.
AE	*Level3*	*Level4* – Acting on the plan.	Creating (source)	Possible negative feedback loop, where the task becomes harder for faster players.

1. The first stage (concrete experience - CE) involves perceiving new knowledge by experiencing the concrete through sensing and being immersed in real situations. In CE stage players will be motivated to explore and discover by collecting gems, a commonly found mineral in volcano lava, in order to fill the gun's tank with water and be able to put out the fires to save the island. The gems and fires are scattered all over the virtual field (island), with the aim of creating *level1* resources (observing): observations of the island. The design includes a positive feedback loop, where the more the player puts out the fires with the water gun, the more he/she has the chance to walk around the island and observe the environmental terrain.
2. The second stage (reflective observation - RO) focuses on watching and reflecting on the learner's own experiences or that of others. In the RO stage, the player will be encouraged to reflect on the observations from the first stage, and *level1* should be converted to *level2* (reflecting) resources. The reflection is performed by finding information about the observed environmental terrain, which is supposed to include observing volcanoes. The player has to collect data about a volcano on the island, such as type, name, and status, and record them in a table inside the game. A positive feedback loop is designed so that if the player finds one piece of information about the observed volcano, he/she can find the next required data.
3. The third stage (abstract conceptualization - AC) is about analysing, thinking, or planning through a symbolic presentation. In the AC stage, the player will experience signs of natural hazards and be in a situation where he/she has to do something to survive. Thus, the player will be prompted to plan by synthesizing, analysing, and hypothesizing from observations (*level1*) and the collected data (*level2*) to survive the natural hazard. The analysis, hypotheses, and plan form *level3* resources. The feedback loop in this stage is optional and the player can choose to reflect again after synthesizing until the final plan is formed. A negative feedback loop could be designed to achieve a balance where it becomes more difficult for the player to hypothesize and plan after forming the initial plan as new signs of natural hazards emerge.
4. The last stage (active experiment - AE) involves doing things. In AE stage, the player will be forced to act on the synthesized survival plan from the third stage and escape the natural hazard. A negative feedback loop could be designed where the task of escape becomes more difficult for players who completed the previous tasks in a shorter amount of time: barriers are added, impeding their route to a vehicle.

CE is the most frequent entrance stage of the ELT cycle (Dieleman & Huisingh, 2006), so the prototyping process designed the VFTG to start with the CE stage. However, the ELT modelling defines the building blocks of each stage, meaning that the game designer/educator can rearrange the building blocks to start the ELT cycle from any stage based on the player's learning style.

6.3.2 The Second Link (Game Modelling)

Establishing the second link of designing the prototype leads to transforming the learning process into gameplay by utilising the matching scheme tool, task model, evaluation model and skill model. The following subsections explain the design process of each step thoroughly.

6.3.2.1 Matching Scheme

The first step to establish the second link is utilising the matching scheme tool to define the VFTG structure and to identify the suitable game elements (see Table 6.2).

Table 6.2: The matching scheme for the prototype.

Pre-	**During**	**Post-**
Start	Concrete experience - CE	Assessment
Adapting : Prior knowledge: Player profile Character: Identity, Choice	Explore: Narrative, Chance, Interaction, Emotion, Multimodal Presentation	Evaluate: Points Provide feedback: Multimodal Presentation Player's acting: Choice, Action Recognition: Badges
Pre-task	Reflective observation - RO	
Tutorial, Instruction, Multimodal Presentation	Reflect: Choice, Interaction, Resources	
Complexity	Abstract conceptualization - AC	
Single location: One-level	Synthesizing: Interaction, Resources, Challenge	
Skill	Active experiment - AE	Progression
Decision making: Time, Choice, Resources	Doing: Challenge, Interaction, Choice, Time, Consequences	Unlock Content

The **pre-phase** consists of four steps: start, pre-task, complexity, and skill. In the start step, the student's *prior knowledge* is informed by the teacher from the school where the experiment was conducted (Belmont Community School), and all the students who participated in the experiment had an intermediate level of comprehensive ability. Learning style is usually not available for secondary school students, meaning it was not utilised in the prototype design. Regarding the character, the Island of Volcanoes is designed to provide a *control of choice* and allow the player to select his/her *identity* as male or female (see Fig. 6.2). The pre-task is designed as a *tutorial* level with *instructions* (see Fig. 6.3) to prepare the player for the tasks along with the general theme of the VFTG being supported by the *multimodal presentation*. The complexity of this prototype is set as one location (*one level*) and long period of time (one ELT cycle). The final step in the pre-phase is identifying the skill, which is the *decision-making skill* that is required in FBL and the workplace in the 21st century, along with the suitable game elements (*time, resources,* and *control of choice)*.

In the **during phase**, *interaction* is utilised to solve the subtasks in all ELT stages where the player has to interact with the environment by exploring and collecting data. *Multimodal presentation* is employed through different forms of learning materials (text and video). The *narrative* is designed to be introduced in the CE stage by an NPC character (Red Dragon) to create an engaging context to act as a hook for the player to go through the following stages. In addition, the *narrative* attaches *emotion* to the learning experience (the danger related to the volcano's eruption and the urge to survive). In the RO stage, the player will be encouraged to reflect on the observation of the existent volcanoes and challenged to learn more about their status (by collecting data). The *control of choice* is applied by giving the player the opportunity to correct any collected data about the volcanoes that is incorrect. *Interaction* is employed in the AC stage by sending signs to the player from the environment and *challenging* them to understand their surroundings to be able to plan an escape and survive. Also, some resources (learning materials/observed information – the behaviour of animals) are utilised to support the process of synthesizing and planning. In the AE stage, a *challenge* is employed where the player has to survive the danger of a volcano within a certain amount of *time* (Fig. 6.14). Also, an element of choice is provided where the player selects the best method (boat or car) to escape the danger. If the player chooses the wrong method to escape, there will be a *consequence*, and the game will continue with *uncertainty* about what will happen as a result of the player's choice. This all depends on the interaction with the environment (observation and learning materials).

The **post-phase** includes a formative assessment with feedback when it is appropriate. The game elements employed are the awarding of *points* for assessment and *badges* for recognition. Feedback is applied according to a *multimodal presentation* – colour and motion(Fig. 6.9/6.10). The player is provided with the *control of choice* to take *action* to find the correct answers and is given two chances. The progression step is designed as *unlocking content (*new tasks, tools, and learning materials).

6.3.2.2 Task Model

The Island of Volcanoes is designed around the mission of surviving by understanding and analysing a natural phenomenon. The player has to survive on the Island of Volcanos by observing, collecting data, planning, and then acting. The task model shapes the process of designing the tasks of the VFTG in three steps: task preparing, task design, and task evaluation where the focus is on further developing the subtasks in the *during* phase from the matching scheme. Table 6.3 summarizes the task model in addition to the detailed explanation.

Table 6.3: Task model of the Island of Volcanoes.

<table>
<tr><th rowspan="2">Task Preparing</th><th colspan="4">Task Design</th><th rowspan="2">Task Evaluation</th></tr>
<tr><th>Game Mechanic</th><th>Game Elements</th><th colspan="2">Learning Outcome</th></tr>
<tr><td>Learning goal</td><td colspan="5">The VFTG will introduce players to the natural hazard (volcanoes).</td></tr>
<tr><td colspan="6">Learning objectives</td></tr>
<tr><td>B_1</td><td>Collecting/putting out</td><td>Interaction chance, narrative, emotion, multimodal presentation</td><td>Identifying</td><td>level1</td><td rowspan="5">In-game assessment</td></tr>
<tr><td>B_2</td><td>Memory mechanic</td><td>Interaction, presentation, choice</td><td>Labelling</td><td>level1</td></tr>
<tr><td>B_3</td><td>Hidden information/infer valuable knowledge</td><td>Interaction, challenge, resources</td><td>Finding & Recording</td><td>level4</td></tr>
<tr><td>B_4</td><td>Categorizing/ classifying</td><td>Interaction, resources, multimodal presentation</td><td>Classifying</td><td>level2</td></tr>
<tr><td>B_5</td><td>Evaluation/ Run</td><td>Challenge, time, interaction, choice, consequence</td><td>Making a decision</td><td>level5</td></tr>
<tr><td>Complexity</td><td colspan="5">One location and long time.</td></tr>
<tr><td>Difficulty</td><td colspan="5">Intermediate</td></tr>
</table>

In the **task preparing step**, the learning elements (goal, objectives, complexity, and difficulty) are defined, as these learning elements are required to develop the subtasks of the *during* phase from the matching scheme (see Section 6.3.2.1). A learning goal is usually a general description of the desired knowledge and skills that students/players need to successfully acquire after a lesson/FT, while learning objectives are a specific description of what exactly students/players are expected to be able to perform by completing a lesson/FT. The learning goal and objectives are based on the learning content selected from the National Curriculum in England for Geography KS3. The learning goal is:

The VFTG will introduce players to the natural hazard (volcanoes).

The learning objectives are:

B_1 = The VFTG will cover the observation of physical features.

B_2 = The VFTG will cover recalling previous knowledge (the structure of the volcano).

B_3 = The VFTG will teach the gathering of geographical information.

B_4 = The VFTG will teach the analysing of natural signs.

B_5 = The VFTG will refine the decision-making skill.

In addition, complexity is identified based on the task variables (time and location) (see Fig. 4.12 in Chapter 4) as a single location and a long period of time. The single-location and long-time variables are suitable to design the subtasks of the ELT stages which give the player enough time to explore the level (island), observe information, and analyse them in addition to developing a high-order skill (decision-making skill). This level of complexity is suitable given the limited time available to design and implement the VFTG, as well as conducting the learning experiment and collecting data. Difficulty in general is defined based on the prior knowledge (students' level of comprehensive ability), which is intermediate. However, each task designed to achieve one of the learning objectives follows the difficulty level based on the revised Bloom's taxonomy.

In the **task design step**, the game mechanics and learning outcomes of each subtask that represent an ELT stage are defined and linked to chosen game elements from the matching scheme. The first two learning objectives (B_1 and B_2) in Table 6.3 are achieved via two subtasks that form the first stage (CE). Two game mechanics are designed to fulfil B_1, which represents the first subtask of the CE stage. The first learning objective (B_1) aims to encourage the player to explore and observe the virtual field by collecting special stones (gems) to fill the gun's tank with water and then put out the fires with a water gun. Both the gems and fires are placed all over the island to guarantee that the player will walk around exploring the natural terrain and discovering the physical features (volcanoes). The game elements applied to achieve the first learning objectives are selected from the *during* phase and CE step of the matching scheme (Table 6.2), while the learning outcome is the ability to identify the number of volcanos in the island as a result of observation. B_2 is accomplished by the second subtask of the CE stage, which is recalling previous knowledge that should be taught to students in geography KS2. A memory mechanic is employed along with interaction and presentation of game elements to achieve the learning objective of recalling the structure of a volcano by labelling its parts, while the learning outcome is that the player will be able to label the structure parts of a volcano.

The third learning objective (B_3) is about teaching the gathering of geographical information in the VFT. Two game mechanics (hidden information and inferring valuable knowledge) are applied to accomplish this learning objective. The player has to find some data (name, country, type, status) about a specific detected volcano after observing the environmental terrain by accessing a learning resource made available via the resources menu (Fig. 6.8). The learning resource is a web page providing information about Bali's volcanoes, which the player can search quickly to find the required data. The web page displays a volcanic map of Bali along with information about each volcano when hovering over them with a mouse, together with the live status of volcanic activities. As well as the web page, text (Fig 6.17) and video learning materials are provided to the player by the Red Dragon (Fig. 6.16), which is an NPC mimicking the role of a leader who provides insights and hints to the player from the start of the VFTG. The learning materials include factual knowledge about possible volcanic activities and the different types of volcanoes. The player can connect the learning materials provided by the Red Dragon to the data found on the web page to understand the level of volcanic activities and the type of selected volcano to infer the situation on the island and predict any possible eruption. This task and associated learning objective represent the RO stage, where the player is expected to reflect on the observation of volcanoes by gathering data to gain further understanding of the situation on the island. The game elements are interaction, challenge, and resources, while the learning outcome is defined as the ability of the player to find and record the requested geographical information.

The fourth learning objective (B_4) concentrates on teaching analysis of the observed natural signs in the VFT. To achieve this learning objective, a task is designed to encourage the player to observe and analyse with the aim of defining the situation on the island. The game mechanic applied is categorizing/classifying the natural signs according to two levels of danger related to the volcano's natural hazards. The player will be presented with one of two different scenarios:

$Scenario1$ = Releasing gas from the active volcano (Agung) and shaking the island (earthquake).

$Scenario2$ = Releasing ash and lava and seeing some animals (deers) running.

The presented natural signs and learning materials in texts provided by the Red Dragon, together with collected data from the previous task, should help the player to synthesise and hypothesise about the situation on the island. The player has to classify the level of danger based on the presented scenario as $Scenario1$ presents less danger level than $Scenario2$. This task and the related learning objective represent the AC stage of ELT modelling. The game elements utilised to design this task are interaction, resources, and multimodal presentation, while the learning outcome identifies the player's ability to classify the natural hazard signs.

The final learning objective (B_5) is to refine the decision-making skill. The task is designed to achieve this learning objective and at the same time to reflect the AE. The player has to evaluate the produced resources (Table 6.1) of all the previous tasks: observing (*Level1*), collecting data (*Level2*), and classifying natural signs (*Level3*). Then the player has to make a decision and act on it (doing), as expected in the last stage of ELT modelling. To survive the natural hazard, the player has to decide to buy a car or boat based on the level of danger. The game elements employed are challenge, time, interaction, choice, and consequence. The time element (Fig. 6.14) is applied to add some pressure and challenge to the VFTG by requiring the player to make a decision in a limited amount of time, with a consequence for making the wrong decision. The learning outcome is that the player's decision-making skills will be improved.

The task model suggested utilising some story tasks with the aim to engage and motivate the player. This VFTG prototype includes some story tasks such as collecting the blue gems. Another story task is exploring the area nearby a volcano that birds are flying over, which leads to accessing the learning task associated with the third learning objective (gathering geographical information). Some of the tasks are introduced as a part the story implicitly by the NPC - Red Dragon (see Fig. 6.16).

In the **task evaluation step**, each identified learning outcome (see Table 6.3) is planned to be allowed to measure by in-game assessment. Also, the assessment tasks should be interweaved into the gameplay and narrative (see Fig. 6.7). The evaluation is designed to be completed while playing the VFTG. For example, the finding and recording of geographical information are designed to be evaluated in the VFTG. The player has to search for some data by accessing a learning resource (Bali's volcanic monitoring web page) through clicking on the exclamation sign (Fig. 6.8) in the resources menu; after finding the required data, the player is expected to record them inside the VFTG. The player's action of recording the data will be evaluated and two coins are awarded for each correct piece of data.

The decision-making skill model (see Chapter 4, Section 4.2.2) was applied to design the last learning task associated with B_5. The next section explains the process of modelling the decision-making skill task, while Section 6.3.2.4 describes the process of identifying the required indicators for each task, building on the resulting design from the task and skill models.

6.3.2.3 Skill Model

The proposed model of the decision-making skill was developed based on simple steps (see Section 4.2.2.4). Each step was mapped to the internal economy mechanic and then translated to learning tasks, as clarified in more detail in Section 6.3.2.4.

Two types of resources are defined in the decision-making skill model as explained in Section 4.2.2.4/Fig. 4.20: *proficiency* and *observation info*. However, it could be difficult to obtain proficiency in decision-making for secondary school students. In view of this, the learning task that reflects the first step (***Identify the problem/situation***) is designed to only employ observational information to identify the problem that needs a decision to be made to solve it. In this prototype, the second step (***Distinguish a list of possible options***) is designed to be easier because the player does not need to come up with the list of possible options; instead, a list (car and boat) is made available, and the player needs to find the list and utilise it to solve the task.

For the third step (***Weigh options accurately and determine priorities***), the model requires the *level3* resource (*synthesizing hypotheses)* to be applied. As shown in Table 6.1, *level3* resource is produced by completing the third stage (AC) of the ELT modelling. The following step (***Select the decision***) was combined with the third step to design one task. The skill model in Section 4.2.2.4 suggests giving the player the possibility of evaluating ("weighing and evaluating" block – Fig. 4.22) the selected decision before executing. However, there was no need to design this step in the prototype because a particular amount of time is specified (risk factor) to make the decision and implement it. Due to this, the player cannot “weigh and evaluate” within the limited time condition.

The fifth step (***Implement the selected decision***) based on the model is expected to produce a *level4* resource – doing (see Fig. 4.23). The last step (***Evaluate the effect of the selected decision and keep a reference for future decision-making situations***) could not be designed as it requires more tasks and levels, which could not be developed within the limited time and resources of this research. As a result, the balancing process (see Fig. 4.24) that is modelled based on employing references to improve the proficiency level and reduce the chance (p3) of selecting the wrong decision could not be designed as it would affect the next decision-making task. However, this prototype is designed to introduce some balance by implementing a risk factor (limited time) that impacts the whole decision-making process (Fig. 6.14).
Some adaptations were made following the decision-making skill model that was presented as a part of the conceptual framework, showing that flexibility is possible for game designers/educators if needed.

6.3.2.4 Evaluation Model

The evaluation model (see Chapter 4) defined the process of assessment in detail as an internal economy mechanic and identified a new and important resource: an indicator. Three types of indicator resources (knowledge, field-based, and skill) were utilised in the process of designing the Island of Volcanoes prototype. The main steps of modelling the evaluation process will be explained:

1. The player solves a task to construct a knowledge/skill:
   When designing the evaluation of tasks in Island of Volcanoes, the concept of solving a task represented a converter that transforms the task and difficulty level to knowledge/skills (see Fig. 4.14). For example, observation should be transformed from the first task of collecting gems and putting out fires based on level1 difficulty as the first level of revised Bloom’s Taxonomy (see Table 6.3). Another example, analysing natural signs to a level of danger by converting the task of classifying the presented $Scenario$ based on level2 difficulty as the second level of revised Bloom’s Taxonomy.

2. The indicator should be identified for tasks.

The learning and game elements that used to identify indicators are distinguished (LG, Bs, KU, Ks, LTs, Ns, As, $M(\,)$) below. The definitions of these elements are explained in Table 4.3.

The Island of Volcanoes prototype contains a small part of the learning content of the KS3 geography curriculum shown in Fig. 6.1. To cover more learning content regarding the same learning goal, the learning objectives could be extended to include different types of natural hazards or to cover learning objectives on human geography such as the effect of natural hazards on human processes (economic activities), or more fieldwork skills such as using maps could be included. The learning goal (LG) is composed of the five previously stated learning objectives (Bs) in the task model (Section 6.3.2.2) plus B_6 regarding human geography, and B_7 regarding utilising maps to detect plate tectonics. Then, LG can be defined based on Eq. 4.1:

$$LG \; = \; [0.18, 0.17, 0.17, 0.17, 0.17, 0.07, 0.07]\; [B_1, B_2\;, B_3,\; , \ldots, B_7]^T$$

This VFTG prototype prioritizes the first five Bs ($B_1, B_2\;, B_3, B_4, B_5$) over the rest of the Bs by assigning them higher weights ($0.18, 0.17, 0.17, 0.17, 0.17$). As explained in the evaluation model (Section 4.2.2.3, Eq. 4.1), the first vector contains weights (w) to signify the importance of each B_i as the values range from 0 to 1 and $\sum w_i$ = 1. B_1 is assigned the largest weight ($w_1 = 0.18$), where the observation of physical features is important for FBL in general. This learning objective is supported by several game mechanics and elements such as multimodal presentation, interaction, and an aerial view of the island. The last two Bs (B_7 and B_7) can be designed to be achieved in the second level or scene in future work.

The knowledge unit (KU) utilised for designing the Island of Volcano prototype is defined (see Eq. 4.2):

$$KU3 \; = \; \{K_1\,, K_2\,, K_3\,, K_4\,,\;\; K_5\;\;, K_6\,, K_7\,, K_8\,, K_9\,, K_{10}, K_{11}, K_{12}\}.$$

The knowledge concepts (Ks) are elicited from the defined learning objectives and stated below:

$$K_1 \; = \; Volcano\text{'}s\ structure\ (crater, lava, secondary\ cone,\ vent,\ magma\ \ chamber,\ layers\ of\ ash\ and\ lava).$$

$$K_2 \; = Items\ can\ be\ used\ \ to\ survive\ natural\ hazard - volcano\text{'}s\ eruption.$$

$K_3 \ = \ Types\ of\ volcano\ activities.$

$K_4 \ = \ Warning\ signs\ of\ volcano's\ eruption.$

$K_5 \ = \ General\ knowledge\ of\ playing\ video\ games.$

$K_6 \ = \ Knowledge\ gained\ from\ observation.$

$K_7 \ = \ Knowledge\ of\ searching.$

$K_8 \ = \ Knowledge\ gained\ from\ searching.$

$K_9 \ = \ Types\ of\ volcanos.$

$K_{10} \ = \ Knowledge\ gained\ from\ analysing.$

$K_{11} \ = \ Knowledge\ about\ plate\ tectonics.$

$K_{12} \ = \ Knowledge\ gained\ from\ keeping\ reference\ of\ decision - making.$

As explained in the proposed framework in Chapter 4, all Ks which belong to one KU have the same level of importance. However, more KUs can be defined in the case of designing the second level to include more Ks that are required to be competent in human geography, plate tectonics, and using maps.

Two pre-requisite KUs ($KU1$ & $KU2$) should be defined based on the KS1 and KS2 geography curriculum. However, these two pre-requisites and the weighted-directed graph of their dependencies will not be defined because the prototype includes one level that utilises $KU3$. The pre-requisites would be employed if the preceding levels were planned to be designed.

The Island of Volcanoes prototype includes two types of learning tasks: practice and assessment. The practice task aims for learning without grading while the assessment task involves grading in addition to learning. The learning tasks of the VFTG prototype are evaluated based on the proposed conceptual framework where the player performs one or more interactions (Ns). The process of evaluating learning tasks is explained below and summarized in Table 6.4 and Table 6.5. The tasks (LTs) of each ELT stage are displayed in Table 6.4 along with the associated learning objectives (Bs), which are defined in the task model (see Section 6.3.2.2). In addition, the required knowledge concepts (K) and actions (As) of each task are shown. The numbering of LTs does not reflect their order of occurrence in the VFTG, but the order of LTs is based on the ELT stages that are associated with them.

Table 6.4: Summary of learning tasks of ELT cycle.

Stage	Objectives	Task	Knowledge Concepts	Actions
CE	B_1	LT_4	K_5	$A2,\ A6$
		LT_5	K_6	$A7$
	B_2	LT_6	K_5 , K_1	$A8,\ A9,\ A4$
RO	B_3	LT_7	K_5 , K_7 , K_8	$A10, A8,\ A4,\ A3,\ A9$
AC	B_4	LT_8	K_5 , K_3, K_9 , K_8, K_4	$A10, A11$
AE	B_5	LT_9	K_5	$A10$
		LT_{10}	K_{10}	$A12$
		LT_{11}	K_{10}	$A13,\ A14$

While Table 6.5 presents the design of the important elements that utilised in the evaluation process based on the evaluation model (see Chapter 4, Section 4.2.2.3) in aim of identifying the required indicators. The LTs and Ks are described while $M()$ is defined, however the order of LTs in this table is not relevant to their occurrence in the VFTG.

Table 6.5: Summary of main descriptions of the evaluation model's elements.

LT	**Description**	**Action**	**Definition**	**K**	**Description**
LT_1	Preparing the survival kit	*A1*	*M(Search,None,Environment,0)*	K_1	Volcano's structure.
LT_2	Reading learning materials	*A2*	*M(Collect,None,Environment,0)*	K_2	Items can be used to survive natural hazards.
LT_3	Watching a video	*A3*	*M(Read,None,Learning Content,0)*	K_3	Types of volcanic activities.
LT_4 LT_5	Observation of physical features	*A4*	*M(Click,None,Environment,0)*	K_4	Warning signs of volcano's eruption).
LT_6	Recalling previous knowledge	*A5*	*M(Watch,None,Learning Content,0)*	K_5	General knowledge of playing video games.
LT_7	Gathering of geographical information	*A6*	*M(Put out, Water Gun, Environment,0)*	K_6	Knowledge gained from observation.
LT_8	Analysing natural signs	*A7*	*M(Write,None,Environment,1)*	K_7	Knowledge gained from searching.
LT_9 LT_{10} LT_{11}	Refining the decision-making skill	*A8*	*M(Walk,None,Environment,0)*	K_8	Knowledge gained from searching.
		A9	*M(Write,None,Environment,0)*	K_9	Types of volcanos.
		A10	*M(Look,None,Environment,0)*	K_{10}	Knowledge gained from analysing.
		A11	*M(Analyse,None,Environment,1)*		
		A12	*M(Select,None,Environment,0)*		
		A13	*M(Ride,None,Environment,0)*		
		A14	*M(Drive,Vehicle,Environment,0)*		

Practice tasks: as explained earlier that practice tasks aims of teaching players some knowledge/skill without grading. The purpose is giving the player/learner an opportunity to practice some tasks without the pressure of assessment. This VFTG prototype includes three practice tasks (Preparing the survival kit, reading learning materials, and watching a video). Their interactions (Ns) and knowledge concepts (Ks) are designed, however the assessment methods $S()$ are ignored.

a) Preparing the survival kit:

This task requires two interactions from the player to be completed and defined based on Eq. 4.6:

$$LT_1 = [0.5, 0.5] [N_1, N_2]^T$$

As explained in the evaluation model (Eq. 4.6), the first vector contains weights to signify the importance of each N_i. The values of w_i range from 0 to 1 and $\sum w_i$ = 1, while the second vector includes the required interactions of this specific task.

$$N_1 = K_2 * A1 + [Kit_1, Kit_2, \ldots, Kit_i] \text{ for } 1 \leq i \leq 5$$

where $A1 = M(Search, None, Environment, 0)$. $N1$ is defined based on Eq. 4.4 as the first interaction needed from the player by utilising $K2$ along with $A1$ on survival kit items (water, food, flashlight, mask, and first aid bag) which can be helpful for escaping a volcanic eruption.

$$N_2 = K_2 * A2 + [Kit_1, Kit_2, \ldots, Kit_i] \text{ for } 1 \leq i \leq 5$$

Where $A2 = M(Collect, None, Environment, 0)$. Both actions ($A1$ and $A2$) employ the same knowledge concept (K_2) and are enforced on the same objects as Kit_i, which represents one of the survival kit items. The $M(\,)$ function (see Eq. 4.11) maps a resource if required and type of interaction ($Environment$) to an action verb (search or collect). In the case of $A1$ and $A2$, there is no need for any resource, which is represented by $None$. The last parameter of $M(\,)$ distinguishes observational action (OA) and a zero value means it is not OA.

A resources menu (see Fig 6.8) is designed to be part of the user interface (UI), where two types of resources need to be collected by the player (survival kit items and coins) and their counts will be displayed. As the player collects the kit's items (Kit_i), the counts of each item will increase and be presented on the resources menu. Even if these items are not used by the player during the VFTG, by completing this task, the player gains confirmation of his/her knowledge (K_2) about which items can be helpful for surviving natural hazards. However, the survival kit items can be utilised to design more learning tasks in future work.

b) Reading learning materials as told by the Red Dragon:

The Red Dragon, which is an NPC, guides the player and tells them about some learning materials such as the types of volcanoes - $Text1$ (composite, shield volcanoes, dome), expected volcanic activities - $Text2$ (active, dormant, extinct), and some warning signs - $Text3$. The player has to read the story, the guidance, and the learning materials shared by the Red Dragon and presented as text (see Fig. 6.17).

$$LT_2 = [0.5, 0.5, 0.5]\,[N_1, N_2, N_3]^T$$

$$N_1 = K_5 * A3 + [Text1]$$

$$N_2 = K_5 * A3 + [Text2]$$

$$N_3 = K_5 * A3 + [Text3]$$

The player performs these three interactions (N_1, N_2, and N_3) at different times while playing the VFTG. Where $A3 = M(Read, None, Learning\ Content, 0)$ and $Text1$, $Text2$, and $Text3$ represent different parts of the displayed learning materials by the Red Dragon. By completing this task, the player will gain knowledge (K_3, K_9, and K_4), which will be utilised in assessment tasks that are designed in this prototype.

c) Watching a video:

The third practice task aims to teach the player about plate tectonics by watching a YouTube video. This task could be designed to be completed via two interactions. The first interaction would involve clicking on a button to start the YouTube video, while the second interaction would require watching the actual video. However, for simplicity and because no assessment method will be attached to this task as the purpose is only learning, the first interaction is sufficient.

$$LT_3 = [0.5, 0.5]\,[N_1, N_2]^T$$

$$N_1 = K_5 * A4 + [Item]$$

Where $A4 = M(Click, None, Environment, 0)$, and $Item$ represents a game item, which is a button. K_5 as previously defined includes any knowledge that is required to play video games such as clicking or shooting.

$$N_2 = K_5 * A5 + [Video]$$

Where $A5 = M(Watch, None, Learning\ Content, 0)$, and $Video$ represents a plate tectonics YouTube video; there is no tool to perform the action, while the interaction type is with the $Learning\ Content$. By completing this task, the player will gain knowledge about plate tectonics (K_{11}), which can be employed to design more tasks in future work.

Assessment tasks: this type of assessment is designed with the aim of evaluating the player's performance**.** After defining each assessment task along with the required Ns and As, the assessment method is explained based on the identified indicators.

a) Task associated with B_1 (observation of physical features):

This task is about exploring the virtual field by collecting special stones (gems) to fill the gun's tank with water and then put out the fires. This task requires two interactions as follows:

$$LT_4 = [0.5, 0.5\ ]\ [N_1, N_2]^T$$

$N1$ is an interaction that is expected to be performed by the player to collect gems.

$$N_1 = K_5 * A2 + [Gem_1, Gem_2, \ldots, Gem_i\ ] \text{ for } 1 \leq i \leq 20$$

Where Gem_i is a special stone that is placed all over the island, while $A2$ is the same action previously defined for collecting survival kit items; there is no need to repeat defining actions when all the parameters of the map function $M(\,)$ are the same.

$$N_2 = K_5 * A6 + [Fire_1, Fire_2, \ldots, Fire_i] \text{ for } 1 \leq i \leq 25$$

where $A6 = M(Put\ out, Water\ Gun, Environment, 0)$, and this action needs a tool in order to be performed, which is the $Water\ Gun$ and it gets filled by N_1. $Fire_i$ represents the fire flames that are distributed everywhere. There are two sizes of flames: the player will gain two scores for each small flame that has been extinguished and three scores for each large flame. By completing this part of the task, the player will gain knowledge by observation (K_6) and the player will need to apply it to perform the second part of this task.

The second part will be unlocked when the score satisfies the condition (≥ 10). The player will be asked to identify the number of volcanoes on the island, and this part of the task involves one interaction:

$$LT_5 = [1]\,[N_1]^T$$

$$N_1 = K_6 * A7 + [Box]$$

Where $A7 = M(Write, None, Environment, 1)$, and Box is a text box. After exploring the island by collecting gems and putting out fires along with taking an aerial view (see Fig. 6.14), the player is expected to observe the natural terrain, which represents K_6 , and then be able to identify the number of volcanoes on the island. $A7$ involves visually recognizing features in the virtual field and for that, the value of the last parameter is set to 1 to reflect observational actions (OA).

The assessment result (AR) of one LT equals the sum of the assessment results of each N multiplied by its weight (the same as the weights attached to Ns that make up the LT) as explained in Eq. 4.7. The assessment result obtained via assessment method $S_i()$ that is designed to fit individual N. For Ns of LT_4, $S()$ can be performed by identifying the minimum and maximum number of gems/fires that the player should collect/put out to verify that they have explored the island as they should have done. LT_5 has one interaction that can be assessed by comparing it to the correct answer.

b) Task associated with B_2 (recalling previous knowledge – the structure of a volcano):

This task will be presented (Fig. 6.9) when the player explores the area near the Agung volcano while exploring the island. Three interactions are required to complete this task:

$$LT_6 = [0.24, 0.38, 0.39]\,[N_1\,, N_2\,, N_3]^T$$

$$N_1 = K_5 * A8 + [Terrian]$$

The first interaction (N_1) involves the action of walking, where $A8 = M(Walk, None, Environment, 0)$, by employing K_5 on the $Terrian$. Every part of the VFTG that can be interacted with by the player is considered as a game object.

$$N_2 = K_1 * A9 + [Label_1, Label_2\,, Label_3]$$

N_2 is performed by $A9$ where $A9 = M(Write, None, Environment, 0)$ via applying $K1$, which involves knowledge of the volcano's structure, on three labels that are selected randomly from a total of six labels. $A9$ differs from $A7$ as it is not an observational action (OA) and the last parameter of $M(\,)$ reflects that by assigning a value of 0.

$$N_3 = K_5 * A4 + [Item_1, Item_2, Item_3]$$

The last interaction (N_3) is needed to give the player the option to receive feedback and then the chance to act on it. $A4$ and K_5 have already been defined and $Item_i$ represents a button that is available for each label (see Fig. 6.9) and the player can perform N_3 after performing N_2 to get feedback (KR). By completing this task, the player will gain confirmation of their previous knowledge (K_1).

N_1 can be assessed by scoring higher for the shortest path but this is not the purpose of this task. However, in this prototype N_1 is not assessed and so the weight is recalculated as below. N_2 is assessed by matching the correct answer and the variation of writing is considered as typing capital or small letters. Also, the assessment can be extended to accept one correct word when the answer consists of more than one word. N_3 can be assessed by scoring the use of a feedback button when needed.

$$LT_6 = [0.0, 0.48, 0.52]\,[N_1, N_2, N_3]^T$$

c) Task associated with B_3 (gathering of geographical information):

This task will be presented (Fig. 6.10) when the player walks near a volcano that some birds are flying over (see Fig. 6.10). The player has to collect at least one item of the survival kit and then the birds will appear flying over one of the three volcanoes randomly.

$$LT_7 = [0.08, 0.08, 0.08, 0.25, 0.25, 0.26]\,[N_1, N_2, N_3, N_4, N_5, N_6]^T$$

$$N_1 = K_5 * A10 + [Birds_1, Birds_2, Birds_3]$$

The first interaction is performed by utilising K_5 on the game objects $Birds_i$, and $A10 = M(Look, None, Environment, 0)$ as the player needs to look for flying birds on the island.

$$N_2 = K_5 * A8 + [Terrain]$$

The following interaction (N_2) is equivalent to N_2 from the previous task. After looking for and finding the flying birds, the player has to walk to the area near the volcano with flying birds. As a result of walking near that volcano, this task will be presented to the player, who will be asked for some information (name, country, type, status) about that volcano.

$$N_3 = K_5 * A4 + [Res]$$

The resources menu (Fig. 6.8) includes two more different types of resources that are made available (volcanic monitoring web page and vehicle) to the player (see Fig. 6.8). N_3 is completed by performing $A4$ and employing K_5 on Res, which represents the exclamation sign on the resource menu. As a result of this interaction, the player will access the volcanic monitoring web page (Fig. 6.11).

$$N_4 = K_7 * A3 + [Page]$$

N_4 is performed by the player to gather the required information displayed to the player after N_2. The player needs to execute $A3$ by applying K_7 on $Page$, where $Page$ represents a specific volcanic monitoring web page that can be accessed via the resource menu. The type of interaction is with $Learning\ Content$ as the monitoring web page provides learning content about the volcanoes on the island of Bali. By completing this part of the task, the player will gain knowledge (K_8) of the required information by searching, which is essential to complete the task.

$$N_5 = K_8 * A9 + [Field_1, Field_2, Field_3, Field_4]$$

After gathering the required information by performing N_4, the player needs to record the data by performing N_5. The player has to execute $A9$ via utilising K_8 on $Field_i$, which represents empty fields (see Fig. 6.10), to record the collected data (name, country, type, status).

$$N_6 = K_5 * A4 + [Item_1, Item_2, Item_3, Item_4]$$

The last N_6 involves the same knowledge concept and action as N_3 from the previous task. The player has the option to receive feedback and act on it, where an $Item_i$ represents a button that is available for each $Field_i$ (see Fig. 6.10).

As mentioned above, the resources menu has two types of resources that need to be collected by the player, one of them being coins. For each correct recorded piece of data in $Field_i$, the player will earn two coins, which will be needed for another task.

N_1 and N_4 can be assessed with technology such as eye-tracking, where both interactions expect the player to perform the action of looking. However, this technology was not available for this research. As a result, N_1 and N_4 are not assessed and the weight is recalculated as below. N_2 is not assessed as it requires walking action from the player and again the purpose of this task is not to evaluate the ability of the player to perform regular game-playing action unless it would affect the final outcome somehow. N_3 is assessed by ensuring that the player accessed the web page by clicking the icon. N_5 and N_6 are assessed by the same methods as the last two interactions of LT_6.

$$LT_7 = [0,0,0.28,0,0.35,0.37]\,[N_1\,,N_2\,,N_3,N_4\,,N_5\,,N_6]^T$$

d) Task associated with B_4 (analysing natural signs):

This task involves presenting one of two scenarios ($Scenario1$ and $Scenario2$) of natural signs that could occur on the island of Bali. The player has to analyse the natural signs and classify the presented scenario according to an appropriate level of danger.

This task requires two interactions to be completed as the player first needs to interact with the environment and the presented natural signs to analyse them and then select the level of danger based on the result of first interaction.

$$LT_8 = [0.50, 0.50]\,[N_1\,, N_2]^T$$

$$N_1 = (K_3, K_9\,, K_8, K_4\,) * A11 + [Sign_1, \ldots, Sign_i] \text{ for } 1 \leq i \leq 3$$

$$A11 = M(Analyse, None, Environment, 1)$$

N_1 requires more than one knowledge concept ($K_3, K_9\,, K_8\,, K_4$) as the player needs as much knowledge as possible to analyse the situation on the island. The player has to perform ($A11$), which leads to analysing the signs ($Signs$) in connection with what has been learnt about the volcanos' activities (K_3), types (K_9), warning signs (K_4), along with the collected data (K_8) to classify the danger level on the island. $Sign_i$ represents two (gas and earthquake as in $Scenario1$) or three (ash, lava, and running deers as in $Scenario2$) natural signs. $A11$ is an observational action where the player has to perform analysis based on observations and relates to previous knowledge concepts and due to that the last parameter of $M(\,)$ is set to value 1 to reflect observational action (OA). By completing this part of the task, the player will gain knowledge (K_{10}) of classifying the situation on the island according to the level of danger.

$$N_2 = K_{10} * A12 + [RadioButton]$$

The player has to select level one (warning of possible eruption) or level two (volcano's eruption). N_2 requires the knowledge (K_{10}) gained from the previous N to perform ($A12$) to select one RadioButton that represents the correct level of danger. Where $A12 = M(Select, None, Environment, 0)$. N_1 requires an analysing action from the player and again is hard to assess yet it can be inferred from N_2. Therefore, the weights are recalculated below.

$$LT_8 = [0, 1] [N_1 , N_2]^T$$

e) Learning task associated with B_5 (refining the decision-making skill):

The decision-making skill model (see Chapter 4, Section 4.2.4) is utilised to identify indicators for this task. The player has to decide to survive the natural hazard presented in terms of natural signs. The decision-making skill model was developed based on simple steps which could represent tasks needed to complete the overall task:

The first step of the decision-making skill model (**Identify the problem/situation**) is achieved by completing the previous task LT_8. In the previous task, the player recognizes a problem, which is the need to survive, and analyses the situation to classify the level of danger. After a few seconds of presenting one of the scenarios ($Scenario1$ or $Scenario2$), a timer (Fig. 6.14) will be displayed on the UI, and this is considered the beginning of this task.

$$LT_9 = [1] [N_1]^T$$

$$N_1 = K_5 * A10 + [Menu]$$

LT_9 includes one interaction and reflects the second step of the decision-making model (**Distinguish a list of possible options**). N_1 is achieved by performing $A10$ on $Menu$ while utilising K_5. The player has to look for a list of possible options, which are the different types of vehicles available to the player, via the resources menu. These items can be purchased by the collected coins gained from completing c) task (gathering geographical information).

Regarding the assessment, N_1 is not assessed as it involves a looking action.

$$LT_{10} = [1]\,[N_1]^T$$

$$N_1 = K_{10} * A12 + [Vehicle]$$

LT_{10} reflects the third and fourth steps (**Weigh options accurately and determine priorities/Select the decision**) as these two steps combined during the designing process of the skill model (see Section 6.3.2.3). After finding the probable options (LT_9), the player has to evaluate these options and identify the benefit and drawback of each, which leads to the selection of the most suitable option ($Vehicle$) based on the level of danger. LT_{10} involves performing N_1 via $A12$ by applying K_{10} , which is the knowledge gained from completing the previous task (classifying the situation on the island to a level of danger) where $A12$ is defined before in LT_8. It is expected that the car will be selected if the player faces $Scenario1$ and the boat if it is $Scenario2.$ In the case of $Scenario1$, the level of danger does not require leaving the island and the player can reach safety by driving the car to an area far from the active volcano. In the case of $Scenario2$, the level of danger forces the player to leave the island by boat to survive the eruption.

N_1 is assessed by comparing the selected vehicle to the correct one regarding the presented scenario.

$$LT_{11} = [0.5\,, 0.5]\,[N_1, N_2]^T$$

$$N_1 = K_5 * A13 + [Vehicle]$$

The next task (LT_{11}) reflects the fifth step of the skill model (**Implement the selected decision**) and produces a *level4* resource (Doing). LT_{11} involves performing two interactions. N_1 is achieved by executing $A13$ and employing K_5 where $A13 = M(Ride, None, Environment, 0)$; the player has to ride the selected $Vehicle$ first.

$$N_2 = K_{10} * A14 + [Place]$$

N_2 is accomplished by performing $A14$ and utilising K_{10} to reach a ($Place$), which is a safe place, using a tool ($Vehicle$). Where $A14 = M(\text{Drive}, Vehicle, Environment, 0)$, and the $Place$ differs based on the presented scenario in LT_8. In $Scenario1$, the safe $Place$ is a far point from the active volcano on the island. The safe $Place$ in $Scenario2$ is any far point in the surrounding area in the sea (see Fig. 6.15).

N_1 can be assessed but it was not necessary as its performance can be indicated by the following interaction and the weights are recalculated. N_2 is assessed by comparing the last place reached by the player to the identified safe places.

$$LT_{11} = [0,1]\,[N_1, N_2]^T$$

The last step of the decision-making skill model is about evaluating the implemented decision and keeping a reference for future situations as this reference is considered as the gained knowledge (K_{12}) from completing this task. As mentioned previously in the task model (see Section 6.3.2.3), this step can be designed as a part of future work. The step can be modelled in the case of designing more than one level where the player can utilise the gained knowledge and developed skill.

Based on Eq. 4.13., the decision-making skill indicator is defined as:

$$SK = \begin{bmatrix} LT_8 \\ LT_9 \\ LT_{10} \\ LT_{11} \end{bmatrix}$$

Proficiency levels (PLS) can be defined based on the Dreyfus model (Dreyfus & Dreyfus, 1980), as in the conceptual framework. However, the task in hand is simple and the PLS is redefined as follows based on Eq. 4.15:

$$PLS = [Novice, Intermediate, Expert]$$

The player's performance can be labelled according to one of the following proficiency levels:

$Novice$: The player could not survive the natural hazard, which may lead to death.

$Intermediate$: The player survived but used the wrong vehicle since if the player selected the boat when $Scenario1$ was presented, the level of danger did not require leaving the island.

$Expert$: The player survived the natural hazard, utilising the correct vehicle for the presented scenario.

The VFTG prototype is designed for a single player, which means that the player (P) indicator is not employed. In addition, the time (T) indicator is utilised for LT_{11}. The completion time ($CT = Te - Ts$) as presented in the proposed framework (see Chapter 4, Eq. 4.8) can influence the assessment result and the requirement to complete the LT_{11} during a specific time interval (two minutes). The completion time (**time_finish**) is saved for three tasks (LT_6, LT_7, and LT_{11}), as explained in the prototype implementation section.

FB indicators include two types: observational and spatial. Observational indicators are identified while defining $M(\ )$ functions (see Eq. 4.11) for each LT by the last parameter as the value 1 distinguishes observational action (OA) and the value 0 otherwise. The spatial indicator can be one of three forms (Eq. 4.12): position, area, or path. This VFTG prototype is designed to employ the path spatial indicator. The player's movement while exploring and solving learning tasks is captured as a series of positions (X_i, Y_i, Z_i) at regular intervals.

3. Appropriate feedback can be generated.
   Feedback is provided for some learning tasks where it is necessary to ensure learning. Feedback is designed by utilising a function $h(\ )$ (see Eq. 4.18), which is defined and explained in Chapter 4 as part of the proposed conceptual framework. The same characteristic of feedback is generated for both tasks in LT_6 and LT_7 which are associated with B_2 and B_3 by $FB1$ as follows:
   $$FB1 = h(result, KR, Visual(colour))$$
   Where KR defines the type of feedback, which is the knowledge of the result. It is designed to be embedded into the environment without providing a report by changing the colour of a button to green when the answer is correct and red otherwise as $Visual(colour)$ identifies the mode of feedback (Fig. 6.9/6.10).
   $$FB2 = h(result, EF, Visual(motion))$$
   $FB2$ represents the feedback that is generated for e) task and associated with B_5. The type of feedback is elaborated feedback (EF) as an arrow points out a safe place when the time for completing the task is running out in motion mode.

4. The player has the choice to reflect on feedback:
   The player has two chances to reflect on the provided feedback for two learning tasks (LT_6 and LT_7), and also can work to improve the displayed score by collecting gems and putting out fires. In addition, the player has the chance to search for and collect the items from the survival kit.

5. Player's reflection on provided feedback affects the scoring rate for the next learning task:
   When the player sees the displayed score on the UI, he/she can reflect by collecting more gems and putting out more fires. This would result in further exploration and observation of the natural terrain, which would affect the score of the next task (identifying) as the probability of answering correctly would increase. Another example is the reflection on provided feedback on gathering geographical information. The player can use two chances if needed and engage in further searching to find the correct data. This would affect the score of the following task, which means that the player would recognise the situation more clearly and proceed with the learning task of analysing natural signs with a deeper understanding.

6. The assigned score affects the difficulty level for the next learning task:

   The level of difficulty increases over time and ELT stages, starting with level1 for the first learning task and reaching level5 in the last task based on revised Bloom's Taxonomy. However, the difficulty of each task could be adapted for each player individually based on his/her score on the previous task. For example, if the player gathered all the geographical information correctly, more natural signs would be presented as part of the following task. Or if the player labelled the parts of the volcano correctly the first time, then the items of the survival kit could be hidden in a cave or in a box that requires a magical key to open, which would lead to an aside quest.

6.3.3 The Third Link (World Modelling)

The world modelling is built based on the three variables defined by the conceptual framework (Chapter 4, Section 4.2.5): space, time, and storytelling. The space of the VFTG is selected to represent the island of Bali (see Fig.6.5/6.6), which contains three volcanoes (Mount Batur, Mount Agung, and Mount Bratan). The time variable is chosen to represent the current time to reflect the recent status of the three volcanos. The task model of the conceptual framework suggests a specific type of narrative for the single-location/long-time VFTG. However, the proposed prototype applies a simple linear narrative to be able to implement it within the available time for this research.

The spatial and temporal representations of Ks utilised in designing LTs are determined by the space and time variables as follows:

- LT_4 and LT_5 (Observation of physical features) are designed based on the place variable, where the player has to explore the island, which is formed regarding the place variable. The spatial representation of K_6 is expected to be gained as a result of completing these tasks.
- LT_7 (Gathering geographical information) is designed based on the place variable, where the player has to collect geographical information about one of the volcanoes, which is defined by the place variable as a part of Bali, and this is a spatial representation of K_8 . Also, this task is designed based on the time variable, where gathering the geographical information is based on the recent monitoring of the volcanoes on the island and this reflects the temporal representation of K_8 .
- LT_8 (Analysis of natural signs) is designed based on the place and time variables, where most of the natural signs were omitted from the only active volcano on the island. That volcano is defined by the place variable and being active is defined by the time variable, which reflects the spatial and temporal representation of K_{10} .
- $LT_9/LT_{10}/LT_{11}$ (Refining the decision-making skill) is also designed based on the place variable, where the safe place is defined as one point of the place variable.

The spatial tracking of the player is captured as a series of positions (X_i, Y_i, Z_i) while playing. The temporal track is also captured for some LTs and saved in the database as the **time_finish** variables and it is explained in the following section.

The story and the tasks are connected to spatial objects: gems and fires, the highest volcano, flying birds, and kit items. The story is utilised to evoke a suitable emotion, which is the feeling of danger that comes with natural hazards and the urge to survive.

To define space and time, their units should be identified as explained in the conceptual framework (Chapter 4, Section 4.2.5). The unit of time is the game time specified for each scene/level. However, given that the VFTG prototype consists of one level, time can be described (Eq. 4.20) as follows:

$$TI = \{Time\ of\ research\}$$

Identifying all the objects that form the space (SP) is not necessary for this prototype but it can be done. However, only the objects that belong to spatial view (has se attribute attached to) are utilised to define the space (Eq. 4.19) as follows:

$$SP = \{Gems, Firs, Highest\ volcano, Flying\ birds, Kit\ items\}$$

The most significant attributes of each object are identified in a vector form. However, the game designer can specify as many attributes as needed. The space objects are defined based on Eq. 4.21 as follows:

$$Gem = [mutable, K_6, SE_2, random\ position]$$

$$Fire = [mutable, K_6, SE_3, random\ position, small/large]$$

$$Highest\ volcano = [fixed, K_1, SE_4, natrual\ position]$$

$$Flying\ bird = [fixed, K_7/K_8, SE_5, random\ position]$$

$$Kit\ items = [mutable, K, SE_6, random\ position]$$

The first element of the attributes vector defines the type of object as $mutable$ or $fixed$. The $fixed$ type object cannot be moved or changed in any way by the player, while the $mutable$ type can be, as is the case with $Gems$, $Fires$, and $Kit\ item$. The following element is the knowledge concept associated with the object. The third element represents a story event that is attached to this object and will be defined below. The last element is the position of the object in the VFTG world. The $random\ position$ means that the object will be replicated and placed in different positions, while the $natrual\ position$ means that the object is placed based on its original position in nature.

The story event is also defined based on its attributes vector (Eq. 4.22) as follows:

$$SE_1 = [Start, Motivation, Introduction]$$

$$SE_2 = [Middle, \text{Engagement}, \text{info}_1]$$

$info_1$ = Blue gem forms deep in the earth until a brew of hot magma pushes it up to the surface. Gemstones are valuable!!!

$$SE_3 = [Middle, Engagement, info_2]$$

$info_2$ = There is fire everywhere, save the island.

$$SE_4 = [Middle,\ interest, info_3]$$

$info_3$ = Discover the area near the highest volcano.

$$SE_5 = [Middle,\ interest, info_4]$$

$info_4$ = Look for flying creatures and don’t forget to check resources.

$$SE_6 = [Middle, challenging, info_5]$$

$info_5$ = Explore the island, you may find things that help you to survive.

$$SE_7 \;= [End, Relieve/disappointment, Result]$$

SE_1 and SE_7 are not attached to any object apart from the Red Dragon and the message bubble. There are other story events, but their attribute vectors are not defined as they are not attached to specific objects. This leads to defining the only level (Eq. 4.23) of this prototype as:

$$S \;=\; TI \;+\; \sum_{i=1}^{5} O_i \;+\; \sum_{i=1}^{11} LT_i$$

6.4 Prototype Implementation

The Island of Volcanoes prototype was implemented via Unity Engine (Unity 2018.2.10f1 (64-bit)). The Unity Engine was selected because of the provision of a free version that can be used for research purposes in addition to the big community of users that supports new users. The implementation follows the detailed design from the previous section. This section will explain the technical parts of the design, in addition to providing some screenshots of the implementation.

The VFTG provides registration for players and character settings to enable the *identity* and *choice* as designed in the matching scheme (see Section 6.3.2.1). The player is allowed (Fig. 6.2) to select their gender (male or female) and to enter their name to show their identity. The following screen (Fig. 6.3) provides options for the player (Play, Help, and Quit). The Play button allows the player to start the VFTG, while the Help button presents the tutorial level that reflects the pre-task step from the pre-phase of the matching scheme.

Figure 6.2: Identity and choice dialogue.

Figure 6.3: Main menu screen.

A heightmap (Fig. 6.4) of digital elevation data of Bali, which is a representation of the ground surface excluding any surface objects, was generated via a web page (terrain.party). The produced heightmap (png image) was converted to a terrain (Fig. 6.5) in Unity 3D by applying available code from the internet. Bali is the space variable and the data that used to generate the terrain is recent which represents the time variable and both variables are defined during the world modelling (see 6.3.3). Basic elements of the environment were added such as water, skybox, simple grass/palms, and the lakes of volcanos (Fig. 6.6). Also, a cycle of day and night was implemented as one minute for morning and 40 seconds for the night.

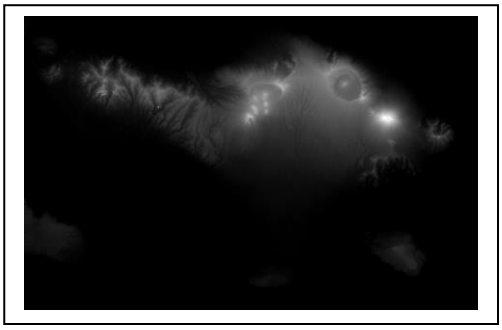


Figure 6.4: Heatmap of Bali.

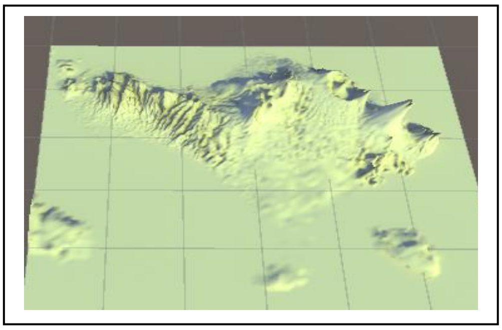


Figure 6.5: Terrain of Bali.

Three icons were added to the top left corner of the (user interface) UI: quests list, resources menu, and settings menu. When the player clicks on the quests list icon (Fig. 6.7), a list of quests (learning tasks) will be displayed as explained in the evaluation step of the task model that the assessment should be interweaved into the gameplay and narrative (Section 6.3.2.2), and a green check will appear next to each quest completed by the player. The resources menu includes two types of resources (Fig. 6.8) that are useful to survive a volcano's eruption in real life. The settings menu provides two options to the player: restart the level and go back to the main menu.

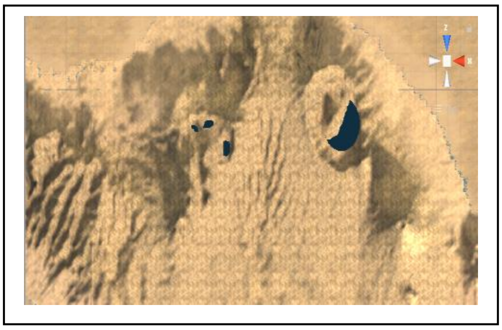


Figure 6.6: Volcanoes' lakes.

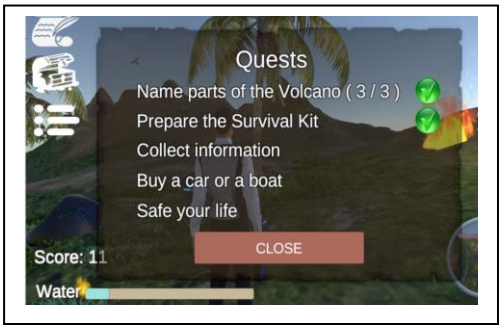


Figure 6.7: Quests list.

Some of the learning tasks (LT_2, LT_3, LT_4, and LT_5) are not displayed on the quests list but are introduced to the player by the NPC (Red Dragon) implicitly as a part of the story (see section 6.3.2.2). For example, the Red Dragon tells the player about blue gems and how they are valuable, which is an indirect way to request the player to perform LT_4 that involves collecting the special stones (Fig. 6.16). Also, some rules were hidden, and the player has to guess them while playing such as the player has to score 10 or more to be able to access the LT_5. Another example of the hidden rules with some hints from the Red Dragon, if the player walks near the Agung volcano, the first quest (LT_6) of the list will appear (Fig. 6.9). However, if the player completes LT_5 and still has not walked near the Agung volcano, the Red Dragon will display a message to encourage the player to do so.

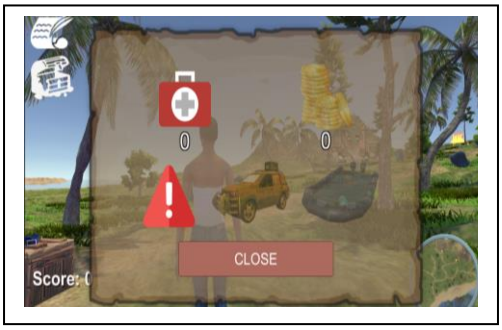


Figure 6.8: Resources menu.

Figure 6.9: First quest (LT_6).

During the exploration, when the player walks near any volcano, its name will be displayed, such as “Mountain Batur”. For both quests: name parts of the volcano (Fig. 6.9) and collect information (Fig. 6.10), there is a check button next to each field to give the player a choice to ask and receive feedback ($FB1$), as discussed in both the matching scheme and evaluation model (see sections 6.3.2.1 & 6.3.2.4). The flying birds appear over one of the three volcanoes randomly after the player collects at least one item from the survival kit. The Red Dragon gives the player hints to explore the area near the volcano where the birds are flying, which leads to displaying the collected information quest (Fig. 6.10), which is LT_6 (gathering geographical information). Fig. 6.11 shows the volcanic monitoring web page that can be accessed from the resources menu by clicking on the exclamation sign (Fig. 6.8). The player can write the first letter of their answers in a capital case or lowercase and also can answer the field of the country in Fig 6.10 as “Indonesia” or “Bali”. The answer to the status field is coded based on all possible statuses of the three volcanoes based on the volcanic monitoring web page at the time of the experiment. The possible statuses are: “minor activity or eruption warning” for the active volcano and “normal or dormant” for the other two volcanoes.

The player gains coins for correct collected data of LT_7 and the count of coins will appear in the resources menu. By completing the quest, natural hazard signs are presented based on the scenarios explained in the evaluation model. Fig. 6.12 shows the gas sign of $Scenario1$ and Fig. 6.13 displays the ash sign from $Scenario2$.

.

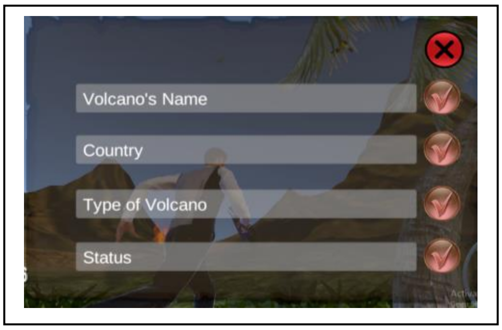


Figure 6.10: Collect information (LT_7).

Figure 6.11: Volcanic monitoring web page.

A timer (Fig. 6.14) is displayed after 30 seconds of selecting the level of danger to give the player a sense of having another quest. The timer creates some pressure on the player to complete the quest within that period of time and also provides a risk factor while they are making the decision as discussed in the process of designing starting from the matching scheme to the evaluation model. If the player has gained some coins (by completing LT_7), then he/she can pay for a vehicle (LT_{10}); otherwise, the vehicle icons in the resources menu will remain inactive. After purchasing the selected vehicle, it appears near the seaside and the player needs to walk or run toward the vehicle to ride it. There are two points (yellow squares in Fig. 6.15) that represent a safe place for $Scenario1$, which are far from the active volcano (Agung).

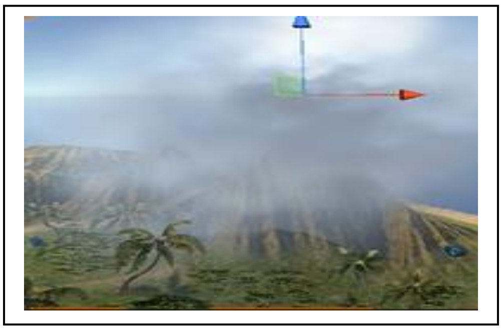


Figure 6.12: The natural hazard sign – gas.

Figure 6.13: The natural hazard sign – ash.

All of the surrounding area (red ellipse in Fig. 6.15) of the sea is a continuous safe point for $Scenario2$. When the time is running out, a green arrow appears and directs towards the safe points as the second feedback provision ($FB2$) that is defined in Section 6.3.2.4. When the player makes the correct decision, a gold badge is awarded and displayed. If the player makes the wrong decision but does not die, a silver badge is awarded. The badge is the game element selected in the evaluation step of the matching scheme tool for recognition.

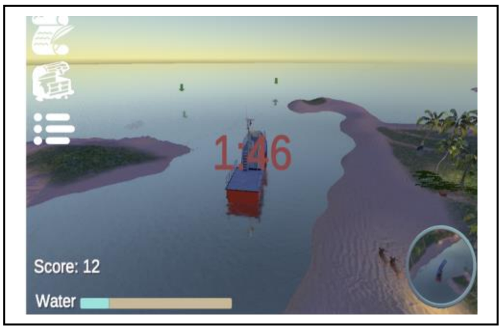


Figure 6.14: The timer of last quest.

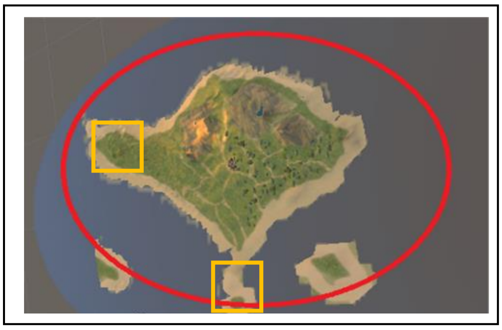


Figure 6.15: Safe points for both scenarios.

The story is told by the Red Dragon - NPC, which is displayed in Fig. 6.16. The story starts by introducing the survival mission to the player by the name that was entered in the character settings (Fig. 6.3). Fig. 6.17. Also, a semi-transparent black image appears as a background while the Red Dragon tells the story.

Figure 6.16: Part of the story.

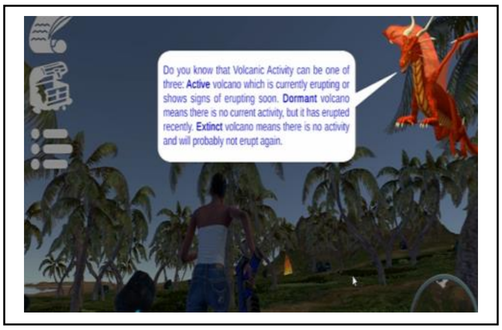


Figure 6.17: Learning materials.

The player avatar and controller were purchased from the Unity asset store, which saved time and effort while implementing the VFTG. The animated Red Dragon character is available for free from several web pages. Also, some pieces of code are available online for free from some forums such as the code for setting the registration and the day/night cycle. The player can control the avatar and the vehicle by arrows or by the (w, a, s, d) keys. A texture of the actual volcano was added to the three volcanoes.

The VFTG was built as WebGL and hosted on a server to be accessible via the internet. Some game data were saved based on the indicators identified in the evaluation model (Section 6.3.2.4). A simple database is utilised for storing the indicators (Fig. 6.18), even for some practice tasks. Saving the game data is achieved in a very basic way after building the database in MySQL. The connection to the database is performed with a server-side script in PHP along with checking the login details to be executed on the server. Game data is passed to the PHP script to be stored in the database via Get request and adding the values of indicators to the URL. Unity offers the UnityWebRequest class to connect with the web server by instantiating a request object. Also, coroutines are used to wait for the request to be completed by using StartCoroutine() method and the yield keyword along with SendWebRequest(). The coroutines momentarily pause and pick up again in the next frame without disrupting the game flow.

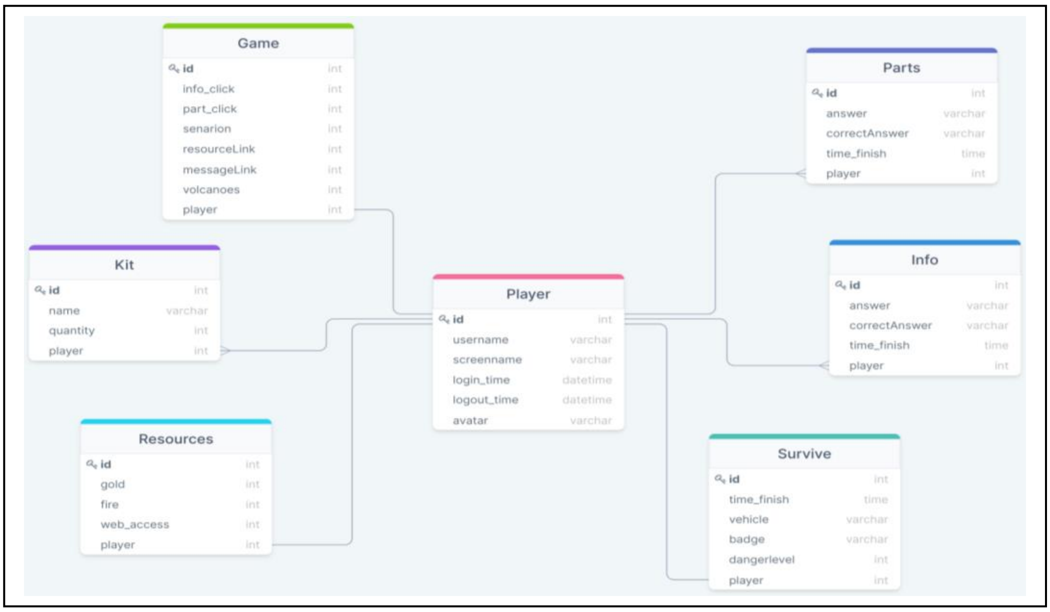


Figure 6.18: Database to save gameplay data.

For each learning task designed in this VFTG prototype, the saved indicators in the database are explained as follows:

- LT_1 (Preparing the survival kit)

  The data for this task was saved in a **Kit** table, which consists of the **name** of the kit item and the **quantity** that the player found and collected (the same item can be collected more than once).

- LT_2 (Reading learning materials as told by the Red Dragon)

  No data are saved on this task. An eye-tracking mechanism while the player reads could be applied to identify some indicators. However, this task was designed only for learning purposes and eye-tracking technology was not available for this research.

- LT_3 (Watching a video)

  If the player clicks on the button that opens the YouTube video, a flag variable is defined to capture the event of clicking on that button and is considered as an indicator of task completion. This flag variable is saved in the **Game** table as **messageLink.**

- LT_4 (Observation of physical features)

  One piece of data is saved for LT_4, which is the number of fires put out by the player, and it is saved in the **Resources** table as **fire**. The number of gems could be also saved but both pieces of data reflect the same thing.

- LT_5 (Identify the number of volcanoes on the island)

  This task is designed to have one piece of data to be saved in the **Game** table as **volcanoes.** This reflects the written value in Box as a result of the player's observation.

- LT_6 (Recalling previous knowledge – volcano's structure)

  The player's input of each $Label$ is saved for both chances in the **Parts** table as the **answer**. The action of clicking the check button to receive feedback is not saved as saving the answers of each chance reflects the same thing. Also, the required time to complete this task is saved in **time_finish** along with the correct answer as **correctanswer** of each $Label$ to show the randomly selected three labels. The number of times that the player opens the task dialogue is saved as **part_click** in the **Game** table.

- LT_7 (Gathering of geographical information)

  The same mechanism is applied as in LT_6; the entry of each $Field$ is saved for both chances in table **Info** as **answer**. Again, the required time to complete this task is saved in **time_finish** along with the correct answer as **correctanswer**. In addition, the action of clicking on the Res, which is the exclamation sign in the resources menu, is saved in the **Game** table as **resourceLink**. The number of times that the player opens the task dialogue is saved as **info_click** in the **Game** table.

- LT_8 (Analysing natural signs)

  The data saved for this task is shown in Table **Survive**, where it is 1 if the player selected level one and 2 if the player selected level two of danger.

- $LT_9/LT_{10}/LT_{11}$ (Refining the decision-making skill)

  Data could be saved for all actions that form a LT. However, actions such as looking ($A10$) for something require a special mechanism. Also, some actions can reflect each other, meaning that saving the data of one action can indicate the other action, such as when reaching the safe place reflects the actions of riding ($A13$) and driving ($A14$) the selected vehicle. Storage limitations should be considered when planning to save data. The data for this task are saved in the **Survive** table as **vehicle** for the selected vehicle and reaching the safe place as a **badge**. If the player reaches the safe place with the correct vehicle in the specified time, a gold badge is awarded. If the player reaches the safe place with the wrong vehicle, which means the player made the wrong decision, the silver badge is awarded. Also, the required time to complete the task is saved as **time_finish**.

General data is saved for each player in the **Player** table. The name of the player is recorded as **name** and **screenname**. Also, the type of avatar (boy or girl) that the player selected is recorded in the **avatar.** In addition, the start time of the playing is recorded as **login_time** and the end time as **logout_time.** The spatial data saved, as mentioned before, is the movement path, which is saved in a text file for each player and named by the player's id.

6.5 Summary

- The VFTG prototype (Island of Volcanoes) was designed and implemented based on the proposed conceptual framework.
- The learning content was selected from the National Curriculum in England for geography KS3 (natural hazard – volcanoes).
- The prototype design created three links: ELT modelling, matching scheme, and world modelling.
- In ELT modelling, the building block of each stage is formed by identifying the required and produced resources, along with possible feedback loops.
- In the matching scheme, the structure of the prototype and the suitable game elements of each block are selected.
- In the task model, the game mechanics and learning outcomes are connected to selected game elements and formed blocks are further developed with clear learning goals and objectives.
- In the evaluation model, the learning objectives of each block along with all designed components (game elements/mechanics, learning outcomes, and assessment) are developed into learning tasks to capture indicators.
- In the skill model, the decision-making skill task is designed.
- In world modelling, three variables are identified (space: Bali, time: research time, story: simple linear story).

- The prototype implementation was done in Unity 3D following the prototype design.
- Some data are saved from the gameplay based on indicators from the evaluation model.

- It is worth mentioning, the proposed framework provides game designers/educators with the flexibility to avoid hit-and-trial behaviour while performing tasks by players. The task model plays a role in guiding the designing process to connect a specific learning objective and its expected outcome to the suitable game mechanic. For example, the game designer/educator can avoid selecting a game mechanic that involves clicking as it is designed in LT_6 and LT_7 where the action *write* is required to solve these tasks instead of *drag and drop* action or *selecting* action via clicking.

Chapter 7

7 Main Study: Data Analysis and Results

7.1 Introduction

The main study of this thesis was a quasi-experiment, which aimed to evaluate the proposed framework by collecting data to prove its actual effectiveness. As the evaluation is essential to prove the effectiveness of designing frameworks, even though some studies skipped it (Feng, González, Amor, Lovreglio, & Cabrera-Guerrero, 2018; Nadolny et al., 2020; Verschueren, Buffel, & Vander Stichele, 2019). However, other studies of frameworks for designing GBL/SG involve at least one form of evaluation such as measuring presence/usefulness (Andreoli et al., 2017), pre/post-assessment (Amengual Alcover, Jaume-i-Capó, & Moyà-Alcover, 2018), functionality/usability problems (Olszewski & Wolbrink, 2017), experts opinions (Roungas & Dalpiaz, 2015), and analysing design factors (Shi & Shih, 2015). This current study adopted more than one form of evaluation which are measuring learning performance and learner experience (presence and motivation) as presented in the following sections. In addition to obtaining experts' perceptions in chapter 5.

The Island of Volcanoes is a VFTG, designed and developed based on the conceptual framework as an evaluation tool. The game is being developed to provide a player-driven experience of natural hazards. It is designed around the mission of surviving by understanding and analysing natural hazards. The learning content is based on the natural hazard topic of volcanoes from the Key Stage 3 geography national curriculum in England. The target audience is secondary school students, and the game world design focuses on Bali, which is an island that contains three volcanoes: Mount Batur, Mount Agung, and Mount Bratan. The player has to survive in the Island of Volcanoes by observing and collecting data.
A quasi-experiment was conducted to determine the learning outcomes of playing the developed game and the efficiency of the conceptual framework. The research ethical approval was obtained from the Ethics Representative of the Department of Computer Science before contacting schools. Also, the enhanced Disclosure and Barring Service (DBS) was gained by the researcher to work directly with children (secondary school

learners) under the supervision of their teachers. All participants are invited to participate in the experiment voluntarily and the teacher contacted the parents to gain their permission.

Hypotheses were analysed to specify how performance achievement differed according to the learning condition. The independent variable was the learning condition (in class or via the VFTG), while the dependent variables were those associated with learning performance. The hypotheses of this experiment are as follows:

The null hypothesis (H_0) is that the learning performance of students who learnt by the traditional method is equal to or less than that of those who learnt by the VFTG method.

The alternative hypothesis (H_1) is that the learning performance of students who learnt by the VFTG method is better than that of those who learnt by the traditional method of learning.

In addition to measuring the learning performance, the study aimed to evaluate the learning experience of the experimental group regarding motivation and presence.

The hypotheses of motivation are:

The null hypothesis (H_0) of motivation measurement is that the VFTG, Island of Volcanoes, does not provide motivation for players.

The alternative hypothesis (H_1) is that the VFTG does provide motivation.

The hypotheses of presence are:

The null hypothesis (H_0) of presence measurement is that the VFTG, Island of Volcanoes, does not provide a sense of presence for players.

The alternative hypothesis (H_1) is that the VFTG does provide a sense of presence.

7.2 Experiment Design

The design of the quasi-experiment is shown in Fig. 7.1. The participants were secondary school students, who were divided into two groups: a control group traditionally taught by the teacher and an experimental group who learnt by playing the VFTG. The duration of the experiment was originally set at one-and-a-half hours, but this was adjusted to be equal to the duration of a class period in English schools, which is around one hour. The class time was divided into four activities: the first activity was an introduction about the experiment and its procedure (10 minutes), the second activity involved completing a pre-test consisting of ten questions (10 minutes), the third and main activity was the learning process (30 minutes): this differed between the two groups with the control group learning in the traditional way and the experimental group learning via the VFTG – Island of Volcanoes, and the last activity included answering two questionnaires IMI/IPQ (10 min). The final step of the experiment was completed four weeks after the main activity (learning) and involved both groups completing a post-test. Two questionnaires were administered to measure motivation and presence regarding the VFTG, which concerned only the experimental group.

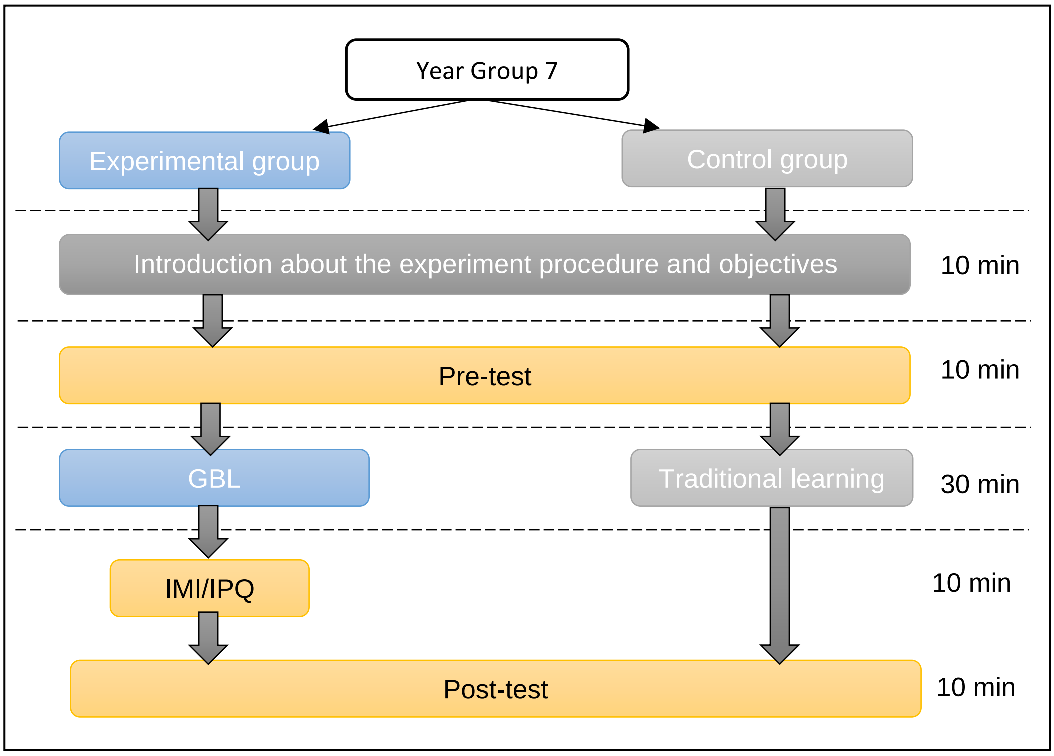


Figure 7.1: Experiment design.

Emails were sent to invite local schools to allow some of their current students to participate in the study. Phone calls were made two days after the emails to explain the nature of the research study to geography teaching leaders. Two schools accepted the study invitation (Belmont Community School and Framwellgate School, Durham) and a meeting was set up to discuss the study with geography teaching leaders. The main researcher visited Framwellgate School and met the geography teacher, Mr Knights, and answered questions regarding the gameplay, learning objectives, and lesson plan. The main researcher also visited Belmont Community School, meeting the head of the geography department, Mr Harvey, and explaining the experiment to him. Discussion of the study and answering questions were followed by emails later. The experiment was scheduled with both schools based on their teaching plans and the availability of PC labs. The ability levels of both the experimental and control groups had to be the same or similar because when measuring the change in gain scores, the change has to be a result of the learning condition (traditional learning or VFTG) and not affected by a difference in ability levels. Both geography teachers agreed to select suitable experimental and control groups that satisfied the condition of similar ability level. However, only one experiment was conducted because the other one, scheduled for May 2020, had to be cancelled due to school closures in the national lockdown. The following is a brief description of the study participants, data collection instruments, and the difficulties faced while conducting the experiment.

7.2.1 Participants

A total of 60 students were divided into two groups: 30 students in the experimental group learnt by playing the VFTG and the other half of the sample were in the control group, where they were traditionally taught by a teacher. The gender of the participants was documented by the students at the start of the experiment. The age of the participants was 11–12 years and they were all in Year 7. The students in the control and experimental groups all had the same level of ability of comprehension.

7.2.2 Instruments

Four instruments were utilised to collect data from the secondary school student participants as follows:

- The first instrument consisted of pre- and post- tests to measure learning performance.
- The second and third instruments were questionnaires that combined together to measure motivation and presence.
- The fourth instrument was the VTFG itself, with log game data collected during gameplay.

The study measures three variables (learning performance, motivation, and presence) because each one of them is essential to ensure effective learning in VFTG. The improvement of learning performance is affected by the player's motivation and feeling of presence as explained in detail in Section 3.4.3. Therefore, the evaluation of the proposed framework lies in the combination of these three variables. This chapter includes a detailed description of each instrument along with information on the data collection and analysis process.

7.2.3 Difficulties

The main study faced some difficulties that meant that some changes were necessary to complete the study. The difficulties and required changes were as follows:

1. The experiment was designed to be completed in more than one hour, which was a problem because the usual time for a class period is exactly one hour. To solve the problem, the number of questions in the pre-test and questionnaire were reduced to allow the whole experiment to fit into a one-hour duration without affecting the game playing time. Originally, the pre-test and post-test were designed to include 15 questions (ten about volcanic hazards and five about Bali's volcanoes), and the motivation questionnaire had four subscales. To solve the problem, the five questions about Bali's volcanoes were deleted and the data collection was confined to the in-game log file in addition to using three subscales in the motivation questionnaire instead of four.

2. An unexpected technical issue occurred when the game did not work on students' PCs but was working on the teacher's PC. It seemed to be the case that the school network was blocking the game web page or that the students' PCs had lower technical specifications. The researcher could not work on the school's PCs to diagnose the issue but tried her best to resolve the issue by emailing the school's technician to identify the problem. WebGL 2.0 seemed to be disabled on students' PCs and cannot be enabled by students. In addition, students' PCs had integrated graphics cards, meaning that the gameplay was very slow. A few changes to the VFTG had to be implemented to adapt the VFTG to the students' PCs.

3. The experiment with Framwellgate Secondary School could not be conducted even after the school opened in the next academic year. The continuation of the COVID-19 pandemic meant that the school denied access to anyone other than students to reduce the possibility of spreading the virus. This difficulty could not be solved.

7.3 Learning Performance

Measuring change in learning performance is a popular method to evaluate interventions in education, with quasi-experiments frequently used for this purpose. The statistical analysis tests selected to play an important role in drawing accurate inferences regarding the significance of an intervention's effects. Pre- and post-test experiments on educational interventions can be evaluated by classic statistical tests such as family-wise analysis of variance (ANOVA) analyses. The simplest statistical test would involve calculating the differences (post-score – pre-score) and conducting a t-test to compare the means of the two groups. This would not take into account “within-test” parameters, such as the difficulty of each item. Moreover, it would not consider the “starting point” of each group, meaning that a difference of 3–1 would be regarded as the same as the difference of 8–6. A more advanced statistical test involves applying a linear regression technique, for example a mixed-effects model. However, this also would not account for each question itself as it would have to be loaded with only the overall score.

Another group of statistical tests are the family of ANOVA statistical tests. These are highly inadequate to analyse educational interventions for a number of reasons: those of relevance to the current research will be explained in the following text. Firstly, ANOVA for use on gain scores, ANCOVA for use on residual scores, and repeated measures ANOVA are all statistical tests used broadly to measure change. They have a strong dependence on statistical assumptions such as homogeneity of variance, normality, independence of observations, and lack of measurement error. However, these assumptions are difficult to satisfy in randomized controlled experiments or quasi-experiments (Alessandri, Zuffianò, & Perinelli, 2017). For example, the intervention elements would produce variations in distributions between groups. Violation of these assumptions can jeopardize the possibility of distinguishing whether differences between the two groups of the experiment, if any, exist due to the intervention. Secondly, depending on raw scores between the pre-test and post-test to measure change has been widely found to produce misleading results. The justification for this is that actual ability, which pre- and post-tests attempt to assess, is not sufficiently represented by raw scores when these scores rely on the level of difficulty of the test items. In fact, a change in raw score does not mean an equal change in ability, which means the relationship between ability and raw scores is non-linear (Dimitrov & Rumrill Jr, 2003). For example, if a student with low ability and another student with higher ability had an equal change in ability scores, this does not mean that both students achieved the same gains from the intervention. In other words, the raw change in ability scores are not interval levels of measurement, so the same variations in raw scores do not represent the same intellectual variations in latent ability.

One of the statistical analysis tests that overcomes the limitations of traditional methods is the linear logistic test model (LLTM), which is a generalization of the Rasch model (RM) (Fischer & Molenaar, 1995). The LLTM can be utilised to measure change under the following assumptions: the unidimensionality of the latent trait being measured, local independence, and fitting the Rasch model (Fischer, 1995). The LLTM differs from the RM in a number of points, one of which being that the difficulty or item parameters are modelled as linear combinations of various basic parameters. In simple words, it breaks down the difficulty parameter into a weighted sum of a set of basic parameters with a total number less than the number of test items. LLTM can be facilitated to measure the change and to define the item difficulty in relation to cognitive operations or other aspects of the item material. The item parameter of a test item at the second time point (*T2*) in a quasi-experiment is calculated as the sum of the item difficulty parameters at the first time point (*T1)* plus the parameters of the change associated with the intervention.

The most appropriate method for this research was the LLTM of the Rasch family for the following reasons: the effects of time and the effect of the intervention could be distinguished, LLTM was suitable for analysing dichotomous responses in the data (correct or incorrect), and the data was collected from the students of one school. If data could have been collected from more than one school, as was planned before the COVID-19 restrictions, then an ANCOVA with multilevel regression (logistic) could have been preferable if its assumptions were met by the data, as this model works well with data where observations are clustered within a second level (school) and first level (students). In addition, linear logistic models with relaxed assumptions (LLRA) could have been applied if there were multidimensional items or if the RM did not fit the data.

7.3.1 Pre-test and post-test

The pre-test and post-test were conducted to measure students' learning outcomes on volcanoes, a natural hazard topic from the geography national curriculum in England (Key Stage 3). The pre-test was distributed to students prior to the commencement of the learning process. After four weeks of the learning intervention, students completed the post-test under the teacher's supervision. Both tests included ten multiple-choice questions that followed the Revised Bloom's Taxonomy, as shown in Table 7.1, where the objective of each item is stated along with its taxonomy. The post-test consisted of the same ten test items in a different order from the pre-test. The pre-test and post-test scores were coded as 1= correct and 0= incorrect with no missing data.

The pre-test was used to determine students' knowledge about volcanoes and natural hazards in general prior to the learning intervention. It was difficult to conduct test piloting because of time and resource limitations. The time schedule for teaching was limited and the teachers could not free more time for the test piloting. However, some steps were taken to overcome this constraint. The first step is about seeking guidance from Dr Nadia Siddiqui, an associate professor in the School of Education, who advised the main researcher to use questions from past exams and to analyse the test for word difficulty. The main researcher collected questions from BBC Bitesize and other websites that provide quizzes for Key Stage 3 geography. However, the questions seemed unprofessional. To resolve this problem, I explored the literature on multiple-choice questions and the Revised Bloom's Taxonomy to edit selected questions based on experts' guidance and standards. The second step involved applying the following improvements:

- Questions covering all learning objectives.
- Questions are based on different items from the Revised Bloom's Taxonomy instead of testing just recall ability.
- Using direct and clearer language for question stems.
- Identifying three options for all the questions instead of using a different number of options for each question individually.
- Making distractors more plausible.

In the third step, the pre-test was analysed by a readability analyser tool called datayze, which confirmed the reading difficulty of this test as suitable for the Year 7/8 English level. While the last step involved seeking feedback about the test from the two teachers who accepted the invitation to participate in the experiment and found the test suitable to the student's knowledge and year group 7.

Table 7.1: Revised Bloom's Taxonomy test instrument.

Objective	***Revised Bloom's Taxonomy***	***Item number***
Explain what a volcano is	Remember	Q1
Recognize types of volcanoes.	Remember	Q5 Q6
Describe its structure (B_2)	Understand	Q2
Understand volcanoes within the context of plate tectonics.	Remember	Q7
Identify ways to predict volcano eruptions in advance (B_3/B_4)	Understand Apply	Q3 Q8
Decision making (B_5)	Evaluate	Q4
Observation (B_1)	Analyse	Q9 Q10

7.3.2 Statistical Analysis

A comprehensive analysis of the main study was conducted to analyse the data collected during the quasi-experiment with the research instruments: pre-test and post-test and questionnaires. Quantitative methods were applied to explore learning outcomes and learning experiences (motivation and presence). The first step involved understanding the pre-test and post-test data (descriptive analysis) and then measuring the change (inferential analysis). The other steps of analysing the questionnaires and possible analysing of game log data are described in detail in other sections.

A descriptive analysis of the data was performed on both the pre- and post-test scores. Two scaled categories of Fail [0–4] and Pass [5–10] were distinguished based on the total score of the ten test items. The number and percentage of students who failed or passed in each group were calculated. The Revised Bloom's Taxonomy is compressed into three categories: Remember, Understand & Apply, and Evaluate & Analyse. This has been done for technical reasons associated with inferential analysis. The total score of each of these categories was calculated for both experimental and control groups in the pre-test and post-test. The pairwise differences in the total score, and for each of the Revised Bloom's Taxonomy categories, were analysed in terms of the median score and interquartile range (IQR). The internal reliability of the pre-test and post-test of each group was confirmed by Cronbach's alpha (α).

In order to analyse the difference in learning performance between the experimental and the control group, a linear logistic test model (LLTM) was implemented. Two time measurement points were applied: the first point was before the educational process commenced and the second point was four weeks after completion. In the two measurement points, each item *i* was considered as a pair of virtual items, a and b, and presented at different time points to the same student υ. Any change of ability θ_υ that happened between the testing points could be explained as a change in the item parameters. The item parameter β_i corresponded to item *i* while β_a^* *and* β_b^* were associated with the virtual items, a and b. As mentioned before, the item parameter in the first measurement point was calculated as $\beta_a^* = \beta_i$ and in the second measurement point as $\beta_b^* = \beta_i + \eta$ where η was the intervention effect. The amount of change $\delta_{\upsilon i}$ for the student υ with the i-th item could then be computed as the difference between the item parameter at *T1* and the later measurement point *T2*, $\delta_{\upsilon i} = \beta_{bi}^* - \beta_{ai}^*$ (Fischer & Molenaar, 1995). The flexibility of LLTM comes from the possibility of linear representation as follows: $\delta_{vi} = w_i^T \eta$ (Mair & Hatzinger, 2007), where w_i is a design matrix (Fig. 7.2). The columns of the design matrix in Fig. 7.2 represent effects ($\eta_{1,}\ \eta_{2,}$ ... η_k) and k represents the number of test items while η_{k+1} refers to the time point, and η_{k+2} refers to group effect. The number of rows of the design matrix equals the production of the number of measurement points by the number of items by the number of groups. The values of the cells show whether or not effects exist.

To apply LLTM for measuring change, the following assumptions need to be satisfied: the regular Rasch model (RM) should hold, and the data should pass some tests to ensure the unidimensionality of test items. The LLTM allocated coefficients to each test item as follows: the time effect (τ) and the intervention effect (η) as logits. The control group estimated the adjusted odds ratio (AOR) of responding correctly with respect to the effect of time only (e^{τ}), while the experimental group considered both the effect of time and intervention ($e^{\tau+\eta}$). Consequently, the independent effect of the

			η_1	η_2	$\dots$	η_k	η_{k+1}	η_{k+2}
Time 1	Group 1	$\beta_1^{*(1)}$	1	0	0	0	0	0
		$\beta_2^{*(1)}$	0	1	0	0	0	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_k^{*(1)}$	1	0	0	1	0	0
	Group 2	$\beta_{k+1}^{*(1)}$	1	0	0	0	0	0
		$\beta_{k+2}^{*(1)}$	0	1	0	0	0	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_{2k}^{*(1)}$	1	0	0	1	0	0
Time 2	Group 1	$\beta_1^{*(2)}$	1	0	0	0	1	0
		$\beta_2^{*(2)}$	0	1	0	0	1	0
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_k^{*(2)}$	1	0	0	1	1	0
	Group 2	$\beta_{k+1}^{*(2)}$	1	0	0	0	1	1
		$\beta_{k+2}^{*(2)}$	0	1	0	0	1	1
		$\vdots$			$\ddots$		$\vdots$	$\vdots$
		$\beta_{2k}^{*(2)}$	1	0	0	1	1	1

Figure 7.2: Design matrix (Mair & Hatzinger, 2007).

intervention could be determined as $AOR = \frac{e^{\tau+\eta}}{e^{\tau}} = e^{\tau+\eta-\tau} = e^{\eta}$. If $\eta = 0$ (no intervention effect), then AOR is $e^0 = 1$, and there is no difference in performance between the two groups. If $\eta = 1$, then $AOR = e^1 \approx 2.73$, meaning that the rate of correct answers in the intervention group will be approximately 2.73 times higher than that of the control.

The application of the LLTM model can distinguish between the effect of time (τ) and the intervention effect (η). The effect of time in the pre- and post-test was analysed by LLTM to clarify each learner's trend of varying performance across time despite the intervention. LLTM was applied to the complete ten test items and the three Bloom's Taxonomy categories: "Remember": four items, "Understand & Apply": three items, and "Evaluate & Analyse": three items.

7.3.3 Results

The results from both the descriptive and inferential statistical tests are presented and explained. The results revealed findings that helped to understand the data, enabling the evaluation of the proposed framework.

Out of a total of 60 students, 32 of the learners (53.3%) were females. The corresponding percentage did not vary across groups (p=0.301). There were 14 females (47%) in the experimental group and 18 females (60%) in the control group. A total of 12 students (40%) in the experimental group scored a "Pass [5–10]" in the pre-test, increasing to 25 students (83.3%) in the post-test. The equivalent numbers in the control group were 12 students (40%) for the pre-test and 20 students (66.7%) for the post-test. The median difference in the total score between the pre-test and post-test was 3 (2–7) for the experimental group and 2 (1–4) for the control group.

Q1 had the highest percentage of correct answers (75%, 90 out of 120 answers), while Q5 had the lowest (40%, 48 out of 120). Detailed descriptive statistics are shown in Table 7.2 for both groups regarding the overall scores and the three Bloom's Taxonomy categories. First, Cronbach's alpha was calculated, giving α= 0.922 for the entire answers, which indicates excellent internal reliability. The control group pre-test and post-test scores were α=0.903 and 0.915, respectively, while the Cronbach's alpha scores for the experimental group pre-test and post-test were α=0.918 and 0.871, respectively.

Table 7.2 presents further descriptive analysis of the overall scores of each group, stating the median and (IQR1–IQR3) for the pre-test and post-test, in addition to the percentage of passes in each case. The counts and percentages of scores (0, 1, 2, 3, or 4) are displayed for the three Bloom's Taxonomy categories (Remember, Understand & Apply, and Evaluate & Analyse) for the pre-test and post-test regarding each group.

As mentioned earlier in this chapter, to measure change using LLTM, RM should hold and the data should pass some tests to ensure the unidimensionality of test items (Mair & Hatzinger, 2007). As a first stage, RM was fitted, and the results are reported in Appendix C; these results are necessary to confirm the accuracy of LLTM results but are not important enough to include in the main body of this research. The RM results indicated that item 5 had the lowest probability of being answered 1 (correctly) in both the pre- and post-test measurements. On the other hand, item 4 had the lowest difficulty of being answered correctly in both the pre- and post-test measurements.

The next step involves assessing the unidimensionality of test items by examining items that do not vary among subgroups. This was done by applying Ponocny's non-parametric T10 for global model fit by means of subgroup invariance and a test to assess constant discrimination of item 1 against a scale comprising all other items (sample n=1000 matrices). Both the null hypothesis of equal item difficulty in both subgroups and equal item discrimination for item 1 vs. the other items could not be rejected. This indicates the RM fitted well enough to the data. To ensure that the RM fits, more than one test of unidimensionality should be conducted, such as Andersen's conditional likelihood ratio test with the mean of raw scores as the partitioning criterion; however, the test displayed a warning message indicating issues related to the majority of items (1, 2, 4, 7, 9, 10) in the post-test estimations and items 3 and 10 in the pre-test estimations. Without these items, the RM fitted the data well according to Anderson's LR test if split by means or median was applied. Thus, an open question emerged about the RM fit where the T10 test indicated that the RM fitted the data well, but Anderson's LR test indicated that the RM fitted well only for several items if split by mean or median was applied. To avoid this issue, random splitting for the LR test was used (all respondents were randomly assigned to two groups) and then the LR test was conducted. The results indicated that the RM fitted the data well. In summary, both Ponocny's non-parametric T10 test and Andersen's LR test indicated that the RM fitted the data well.

LLRM could then be applied with the assurance of accurate results. The LLTM function in R took as input the question answers, a data matrix (60 by 20), and automatically generated the design matrix w. For ten questions, given twice (pre- and post-test), the design matrix consisted of 40 rows and 11 columns, as displayed in Fig. 7.3.

Table 7.2: Descriptive summary of test scores.

	Experimental		Control	
	Pre-test	Post-test	Pre-test	Post-test
Gender				
Female	14 (47%)		18 (60%)	
Male	16 (53%)		12 (40%)	
Overall score				
Scaled from 0 to 10	2 (1–7.5)	9 (5.75–10)	3 (1–8)	7.5 (3–10)
Pass [5–10]	12 (40%)	25 (83.3%)	12 (40%)	20 (66.7%)
"Remember" score				
0	9 (30%)	2 (6.67%)	8 (26.67%)	1 (3.33%)
1	8 (26.67%)	6 (20%)	6 (20%)	9 (30%)
2	4 (13.33%)	2 (6.67%)	8 (26.67%)	5 (16.67%)
3	5 (16.67%)	6 (20%)	5 (16.67%)	2 (6.67%)
4	4 (13.33%)	14 (46.67%)	3 (10%)	13 (43.33%)
"Understand & Apply" score				
0	14 (46.67%)	4 (13.33%)	13 (43.33%)	7 (23.33%)
1	4 (13.33%)	2 (6.67%)	6 (20%)	8 (26.67%)
2	7 (23.33%)	5 (16.67%)	3 (10%)	1 (3.33%)
3	5 (16.67%)	19 (63.33%)	8 (26.67%)	14 (46.67%)
"Evaluate & Analyse" score				
0	13 (43.33%)	0	10 (33.33%)	5 (16.67%)
1	7 (23.33%)	2 (6.67%)	8 (26.67%)	3 (10%)
2	2 (6.67%)	3 (10%)	1 (3.33%)	2 (6.67%)
3	8 (26.67%)	25 (83.33%)	11 (36.67%)	20 (66.67%)

The 40 rows represent each question (10 pre-/control, 10 pre-/intervention, 10 post-/control and 10 post-/intervention), while the 11 columns represent the effects. In particular, the first nine columns (eta1, eta2 …, eta9) represent the effects associated with each question's inherent difficulty while the item parameter for first question had to be fixed to 0 as baseline difficulty. The 10th column (eta10) shows the time effect: all questions that were in the post-test (last 20) have the numeral 1 in the cells of these columns and the numeral 0 otherwise. The last column (eta11) reflects the intervention effect, where the numeral 1 appears for questions from the post-test of the intervention group.

The LLTM model showed the statistically significant positive effect of time, with $\tau = 2.112$ for the overall score scale with ten items (p<0.001). Therefore, the adjusted odds for succeeding in the post-test against the pre-test, without the intervention effect, were $AOR = e^{2.112} = 8.265$ (95% CI = 4.681, 14.591). Moreover, the effect of the intervention was also found to be significant, with $\eta = 1.704$ (p<0.001), indicating an improvement in the overall score of the post-test results from the intervention.

	eta1	eta2	eta3	eta4	eta5	eta6	eta7	eta8	eta9	eta10	eta11
preQ1 t1 g1	0	0	0	0	0	0	0	0	0	0	0
preQ2 t1 g1	1	0	0	0	0	0	0	0	0	0	0
preQ3 t1 g1	0	1	0	0	0	0	0	0	0	0	0
preQ4 t1 g1	0	0	1	0	0	0	0	0	0	0	0
preQ5 t1 g1	0	0	0	1	0	0	0	0	0	0	0
preQ6 t1 g1	0	0	0	0	1	0	0	0	0	0	0
preQ7 t1 g1	0	0	0	0	0	1	0	0	0	0	0
preQ8 t1 g1	0	0	0	0	0	0	1	0	0	0	0
preQ9 t1 g1	0	0	0	0	0	0	0	1	0	0	0
preQ10 t1 g1	0	0	0	0	0	0	0	0	1	0	0
preQ1 t1 g2	0	0	0	0	0	0	0	0	0	0	0
preQ2 t1 g2	1	0	0	0	0	0	0	0	0	0	0
preQ3 t1 g2	0	1	0	0	0	0	0	0	0	0	0
preQ4 t1 g2	0	0	1	0	0	0	0	0	0	0	0
preQ5 t1 g2	0	0	0	1	0	0	0	0	0	0	0
preQ6 t1 g2	0	0	0	0	1	0	0	0	0	0	0
preQ7 t1 g2	0	0	0	0	0	1	0	0	0	0	0
preQ8 t1 g2	0	0	0	0	0	0	1	0	0	0	0
preQ9 t1 g2	0	0	0	0	0	0	0	1	0	0	0
preQ10 t1 g2	0	0	0	0	0	0	0	0	1	0	0
postQ1 t2 g1	0	0	0	0	0	0	0	0	0	1	0
postQ2 t2 g1	1	0	0	0	0	0	0	0	0	1	0
postQ3 t2 g1	0	1	0	0	0	0	0	0	0	1	0
postQ4 t2 g1	0	0	1	0	0	0	0	0	0	1	0
postQ5 t2 g1	0	0	0	1	0	0	0	0	0	1	0
postQ6 t2 g1	0	0	0	0	1	0	0	0	0	1	0
postQ7 t2 g1	0	0	0	0	0	1	0	0	0	1	0
postQ8 t2 g1	0	0	0	0	0	0	1	0	0	1	0
postQ9 t2 g1	0	0	0	0	0	0	0	1	0	1	0
postQ10 t2 g1	0	0	0	0	0	0	0	0	1	1	0
postQ1 t2 g2	0	0	0	0	0	0	0	0	0	1	1
postQ2 t2 g2	1	0	0	0	0	0	0	0	0	1	1
postQ3 t2 g2	0	1	0	0	0	0	0	0	0	1	1
postQ4 t2 g2	0	0	1	0	0	0	0	0	0	1	1
postQ5 t2 g2	0	0	0	1	0	0	0	0	0	1	1
postQ6 t2 g2	0	0	0	0	1	0	0	0	0	1	1
postQ7 t2 g2	0	0	0	0	0	1	0	0	0	1	1
postQ8 t2 g2	0	0	0	0	0	0	1	0	0	1	1
postQ9 t2 g2	0	0	0	0	0	0	0	1	0	1	1
postQ10 t2 g2	0	0	0	0	0	0	0	0	1	1	1

Figure 7.3: Automatically generated design matrix.

This effect corresponds to an $AOR = e^{1.704} = 5.496$ (95% CI = 2.248, 13.434) improvement in the post-test answers for the experimental group against the control group.

The LLTM model showed the statistically significant positive effect of time, with $\tau = 2.112$ for the overall score scale with ten items (p<0.001). Therefore, the adjusted odds for succeeding in the post-test against the pre-test, without the intervention effect, were $AOR = e^{2.112} = 8.265$ (95% CI = 4.681, 14.591). Moreover, the effect of the intervention was also found to be significant, with $\eta = 1.704$ (p<0.001), indicating an improvement in the overall score of the post-test results from the intervention. This effect corresponds to an $AOR = e^{1.704} = 5.496$ (95% CI = 2.248, 13.434) improvement in the post-test answers for the experimental group against the control group. The significance of the time effect (higher scores in post-testing compared with pre-testing) was also retained for the three Bloom's Taxonomy categories of questions: "Remember", "Understand & Apply", and "Evaluate & Analyse", with τ equal to 1.853, 1.872, and 3.603, respectively (p<0.001, for all sub-categories). The impact of the intervention effect on the "Understand & Apply" and "Evaluate & Analyse" score scales was significant, with η equal to 2.5 (p=0.017) and 2.2 (p=0.04), respectively. On the other hand, the intervention did not affect the "Remember" score, with η = 0.843 (p=0.085). Table 7.3 summarises these LLTM results in an organised manner, showing the increased performance across time as well as the impact of the intervention on the overall score in addition to the three Bloom's Taxonomy categories.

Table 7.3: LLTM results for time and intervention effects.

Score	$\boldsymbol{\tau}$	$\boldsymbol{\eta}$
Overall	2.112 ** (SE=0.29, p<0.001)	1.704 ** (SE=0.456, p<0.001)
Remember	1.853 ** (SE=0.415, p<0.001)	0.843 (SE=0.614, p=0.085)
Understand & Apply	1.872 ** (SE=0.559, p<0.001)	2.5 * (SE=1.177, p=0.017)
Evaluate & Analyse	3.603 ** (SE=0.784, p<0.001)	2.2 * (SE=1.255, p=0.04)

Note: All values are described as logits with standard errors (SE) and p-values (p). * indicates significance at p=0.05 level. ** indicates significance at p=0.01 level.

The average overall scores for both groups at the two measurement points are plotted in Fig. 7.4; a positive time effect is indicated for the two groups. However, the slope of the intervention group line is steeper compared to the slope of the control group, representing a positive intervention effect over the traditional learning associated with the control group.

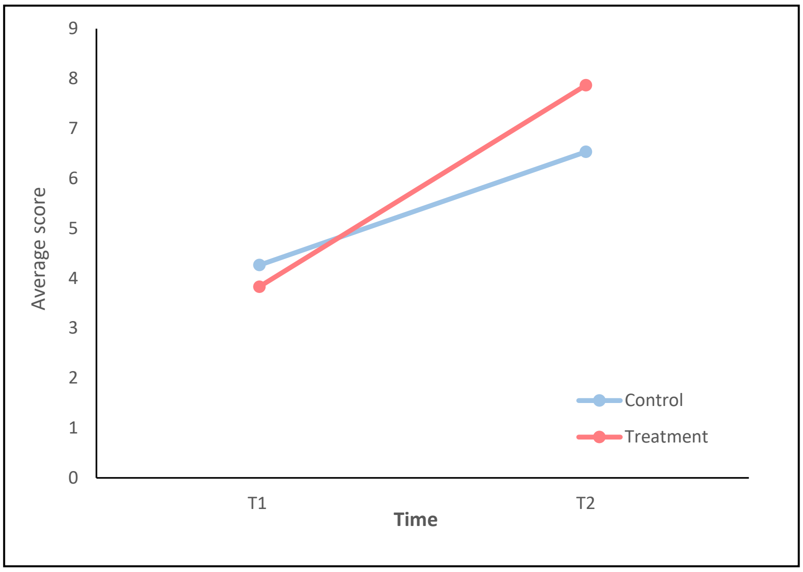


Figure 7.4: Changes in average scores across groups.

The subfigures in Fig. 7.5 present the scores of both groups (control in blue and experimental in red) in boxplot format, enabling a comparison of the pre-test and post-test scores. The change in median scores and IQRs confirms the LLTM results. The subfigures are labelled as followed: a) Overall score, b) “Remember” category, c) “Understand & Apply” category, and d) “Evaluate and Analyse” category.

7.3.4 Findings

The null hypothesis (H_0) is that the learning performance of students who learnt by the traditional method is equal to or less than that of those who learnt by the VFTG method. The alternative hypothesis (H_1) is that the learning performance of students who learnt by the VFTG method is better than that of those who learnt by the traditional method of learning. The descriptive information of students (participants) in both groups (control and experiment) helped to test the hypotheses. All students had the same level of comprehensive ability, which is reflected in the equal pass percentage in the pre-test in both groups at 40%. In addition, all participants belonged to the same age range in Year 7 of secondary school, and the percentage of female and male students did not vary across the two groups. This descriptive information established a similarity of participants that provided control over some variables (ability, age, and percentage of females and males) so that they would not impact the learning performance. The descriptive analysis shows that Q1 had the highest percentage of correct answers (75%), Q4 had the second highest (73%), while the lowest percentage was in Q5 (40%). This matched the results of RM where Q5/item 5 had the lowest probability to be answered 1 (correctly) in both the pre-test and post-test, while Q4/item 4 had the lowest difficulty to be answered correctly at both measurement points. However, Q1/item 1 was not included in the RM results because it was the baseline with a difficulty estimation equal to 0. The main point of analysing test items was to establish their reliability and validity as part of the data collection instrument and to understand their difficulties and effects on learning performance. LLTM was selected to determine the test items' effects on learning performance and there was no need to analyse these items further.

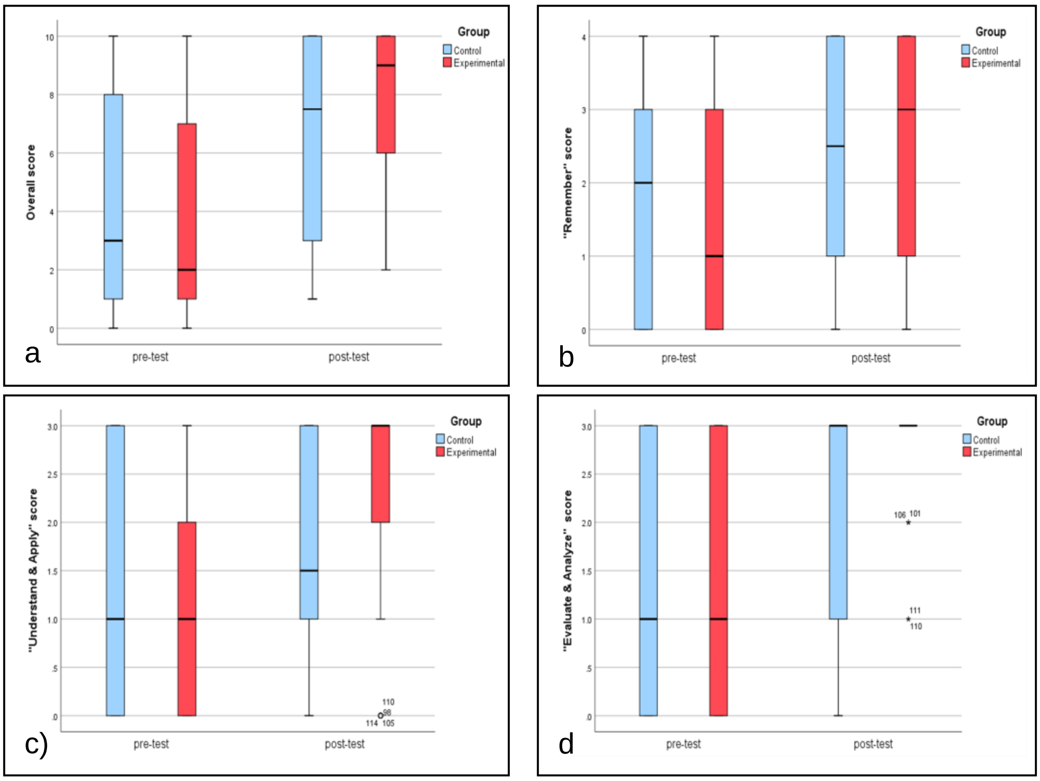


Figure 7.5: Boxplots for the score of both groups, both pre- and post-test on a) Overall score, b) "Remember" category, c) "Understand & Apply" category, and d) "Evaluate and Analyse" category.

The main finding was that the educational intervention, the VFTG, improved learning performance to a greater extent than the traditional method of learning. Traditional learning in this study took place through class learning with traditional paper exercises and images on PowerPoint slides. The VFTG method enhanced the "Understand & Apply" skills along with the "Evaluate & Analyse" skills compared with the traditional method of learning. However, the “Remember” skill stayed the same in the two groups. In summary, based on LLTM we can reject the null hypothesis.

The VFTG method showed promising effects on learning performance in a topic that should be taught in a field trip (FT) but cannot be for different reasons facing teachers around the world, the main reason being student safety. Natural hazards such as volcanoes are not usually taught in a secondary school FT. The traditional method of learning is the only way to teach this topic, using some technologies such as video clips which have limited applicability to learning theories and the sense of being there, among other issues discussed in the literature review chapter. The VFTG provides an opportunity of having a FT when an actual FT is impossible to deliver. The main assumption identified at the beginning of this research is that the VFTG method is not a replacement method for actual FTs, but is an alternative method. Therefore, the study did not compare the learning performance of a VFTG with that of an actual FT; also, it would be difficult to send secondary school students on a FT near an active volcano to conduct an experiment.

7.4 Motivation

Motivation is an essential element in many aspects of life such as sports, learning, and work. It can be defined broadly as a force that stimulates and regulates behaviour or desire. Two types of motivation can be distinguished: intrinsic and extrinsic motivation (L. C. Wang & Chen, 2010). Intrinsic motivation has been considered as conducting a task for the ingrained pleasure and satisfaction of the task itself (Ryan & Deci, 2000), while the opposite is extrinsic motivation empowered by external rewards such as money, marks, or praise resulting from pursuing and engaging in a task (Moos & Marroquin, 2010). Previous studies considered a number of different terms referring to motivation such as flow (L. C. Wang & Chen, 2010), engagement, or even immersion (Garneli, Giannakos, & Chorianopoulos, 2017). In general, these terms are used interchangeably, but sometimes they are viewed as related constructs (Pesare, Roselli, Corriero, & Rossano, 2016).

Intrinsic motivation is more relevant to playing digital games, which are believed to be intrinsically motivating activities. Many SGs, educational games, or GBL need to be designed to produce intrinsic motivation as it is linked to increased learning (Malone, 1981). The overall purpose of SGs and VFTGs is the achievement of learning (Denis & Jouvelot, 2005), which should be motivated by the elements and mechanisation of play. Measuring intrinsic motivation is one of many factors utilised to evaluate SGs and

GBL (Petri, von Wangenheim, & Borgatto, 2016). This study measured the intrinsic motivation of the VFTG prototype with the aim to evaluate the proposed conceptual framework along with learning performance and presence.

Reviewing the literature revealed two main methods for measuring intrinsic motivation: subjective and objective (Derbali & Frasson, 2012). Subjective methods apply self-report questionnaires, interviews (Hoffman & Nadelson, 2010), and observations (Kremer, 2012). Questionnaires are the most used tools for measuring intrinsic motivation (Coenen, Mostmans, & Naessens, 2013; Hauge, Barenbrock, & Thoben, 2017; Parchman, Ellis, Christinaz, & Vogel, 2000). The questionnaire method can be criticized if the questionnaire validity is not approved by, or created based on, theoretical theories or frameworks. In addition, some studies have stated that this method could capture only the last emotion of the player before the questionnaire was filled in. A small number of studies have utilised structured observation to collect evidence of motivation such as happiness (smiling) or distraction (sadness/unaroused) (Kremer, 2012). This method could be a helpful tool for commercial video game designers, but SGs require a more scientific measurement tool (Fowler, 2013). In my opinion, observation demands a longer time and more human resources to observe the research participants. The same issue applies to the interview method, which in the current study would involve conducting interviews with 30 school student participants; however, they would be unable to skip their classes to take part in interviews, and it would also place demands on the time available for conducting the experiment.

On the other hand, objective methods rely on interpreting physiological (Derbali & Frasson, 2012) or interaction data (Hershkovitz & Nachmias, 2008). External equipment such as sensors is required to capture physiological data. Research studies have utilised different physiological information such as eye-tracking (Ghergulescu & Muntean, 2012), heart rate, skin conductance, or electroencephalography (Derbali & Frasson, 2012) to evaluate intrinsic motivation. Objective methods are promising where the collected data are not precisely dependent on a learner’s opinion. In addition, collecting data by objective methods takes place in real time, and is more accurate (objective) for assessing intrinsic motivation. However, this objectivity involves controlled experiments with sufficient sample size and also requires experience in electrical signal processing and detailed monitoring of changes in these recorded signals. Collecting interaction data is more practical and at the same time provides objective assessment. Interaction data are gathered while the player interacts with the SG, and capture behaviour and interaction. Examples of interaction data obtained to evaluate intrinsic motivation are error rate (Miljanovic & Bradbury, 2018), number of completed tasks (Farrell & Moffat, 2014), and time spent on the game (Ben-Zadok, Leiba, & Nachmias, 2011).

Despite the potential of physiological methods, the current research applied a self-report questionnaire. Utilising sensors to obtain physiological data from school students or even undergraduate students would be too costly in terms of time and resources, especially given the social distance rules enforced because of COVID-19 such as lab seating experimental requirements. Moreover, it would require specific knowledge of electrical signal processing, which is out of the scope of this research.

The Intrinsic Motivation Inventory (IMI) was chosen as a subjective tool because of its characteristics such as psychometric properties, which are reliable and valid (Tsigilis & Theodosiou, 2003), and its feasibility where researchers can select suitable items from the questionnaire's subscale items, adjusting the questionnaire length and the concepts to measure. Also, IMI is built on well-established theory (Self-determinism Theory, SDT; (Ryan & Deci, 2000), and has been applied in a number of SGs and GBL studies (Burke et al., 2010; Eseryel et al., 2014; Hauge et al., 2017; Nieuwhof-Leppink, de Jong, van de Putte, & Schappin, 2019; Vandercruysse, Vandewaetere, Cornillie, & Clarebout, 2013). The following is a detailed explanation of the statistical analysis, together with the IMI results.

7.4.1 Motivation Questionnaire

The Intrinsic Motivation Inventory (IMI) was adapted to measure the students' subjective experience regarding the Island of Volcanoes game. The IMI (McAuley et al., 1989; Plant & Ryan, 1985) is a tool to evaluate motivation based on seven subscales (interest/enjoyment, perceived competence, value/usefulness, felt pressure and tension, value/usefulness, perceived choice, and relatedness). Only the relevant subscales to this research were selected. The adapted version of the IMI employed in this research consisted of 19 statements scored on a seven-point Likert scale from not at all true (1) to very true (7). The selected subscales were: Interest/Enjoyment (7 statements), Pressure/Tension (5 statements), and Value/Usefulness (7 statements), and the statements could be ordered randomly regardless of the subscales, as shown in Table 7.4 where "Statement No." determines the order of each statement in the distributed questionnaire. In addition to the IMI statements, learners were asked to identify their gender and to state how often they played entertaining video games per week. The scale for the frequency of playing video games consisted of five options: every day, 3–5 times per week, 1–2 times per week, not very often, and not at all.

The IMI is an adaptable measurement tool that defines a person's level of intrinsic motivation related to a specific activity such as playing a game. The tool permitted exploration of the effects of the VFTG and provided evidence on intrinsic motivation at a deep meaningful level by analysing variables assumed to facilitate an intrinsically motivated attitude. Measuring motivation, in addition to the feeling of presence in the VFTG, contributed to the evaluation process of the conceptual framework employed to design and implement the Island of Volcanoes game.

Table 7.4: Adapted IMI statements.

Statement No.	Subscale
Interest/Enjoyment	
S1	I enjoyed playing this game very much.
S4	This game was fun to play.
S5	I thought this was a boring game. (R)
S8	This game did not hold my attention at all. (R)
S10	I would describe this game as very interesting.
S14	I thought this game was quite enjoyable.
S16	While I was playing this game, I was thinking about how much I enjoyed it.
Pressure/Tension	
S2	I did not feel nervous at all while playing this game. (R)
S6	I felt very tense while playing this game.
S11	I was very relaxed playing this game. (R)
S13	I was anxious while playing this game.
S18	I felt pressured while playing this game.
Value/Usefulness	
S3	I believe this game could be of some value to me.
S7	I think that playing this game is useful for learning about volcanoes.
S9	I think this is important to play because it can encourage learning more about volcanoes.
S12	I would be willing to do this again because it has some value to me.
S17	I think playing this game could help me to improve my knowledge about volcanoes.
S19	I believe playing this game could be beneficial to me.
S15	I think this is an important game.

Table 7.4 displays statements for each subscale along with the order in which they were applied to the distributed questionnaire. The symbol (R) next to a statement indicates a reverse meaning to the rest of the statements in the

subscale. The questionnaire was distributed to learners immediately after playing the game.

7.4.2 Statistical Analysis

The first step in the analysis process involved calculating a score for each subscale. The reverse statements, which are indicated by an (R) displayed next to them in Table 7.4, were scored by subtracting the statement response from 8, and considering the result as the statement score. The scores of the remainder statements were equal to the response (1 to 7). The subscale score was calculated by computing the mean of the statement scores on the subscale. Additional variables were analysed in this experiment: total post-test score, difference (post–pre) in test scores, gender of learners, and frequency of playing. Scores of the three IMI subscales are described by medians and IQR. The gender variable included two distinct levels while the frequency of playing games had five distinct levels, and these two variables are described in frequencies and percentages. The Cronbach's alpha (α) was obtained for each subscale as a measure of internal consistency.

The scores of the subscales were tested for pairwise correlations with Spearman's rho (ρ) correlation coefficient. Each variable was analysed for association with each one of the IMI subscale scores. The Mann-Whitney U test for independent samples was used to discover the effect of gender on the subscale scores, while Spearman's rho (ρ) correlation coefficient was applied for the rest of the variables.

7.4.3 Results

Two levels of analysis were applied to the collected data from the adapted IMI questionnaire: descriptive and inferential statistical analysis. The descriptive analysis summarizes the data and measures the central tendency while inferential analysis helps to draw conclusions about the data, providing necessary insights into the effectiveness of the conceptual framework. The significance level was set to 0.05 for all comparisons. All statistical analyses were performed in IBM SPSS Statistics (version 26 for Windows). A total number of 30 students were assigned to the experimental group, 14 females and 16 males: all of them answered the questionnaire. Cronbach's alpha was computed for the subscales: Pressure/Tension (5 items, α =0.946), Interest/Enjoyment (7 items, α =0.968), and Value/Usefulness (7 items, α =0.916); the results indicated high internal consistency and reliability for the statements of each IMI subscale.

Table 7.5 provides descriptive information on the variables (total post-test score, difference score, gender of learners, and frequency of playing) which is the frequency or median (IQR). The median of the total post-test scores was 9 (IQR = 5.75–10), and the median of the difference scores equalled 3 (IQR = 2–7). The frequency of playing entertaining video games was recorded as follows: 12 students (40%) reported playing games every day, 10 (33.3%) played three to five times per week, and 3 (10%) played one to two times per week. A small number of students did not play video games as much as would be assumed from this generation: three students (10%) said “not very often”, and only two (6.7%) claimed not to play video games at all.

Table 7.5: Variable characteristics.

Variable	*Frequency (%) or median (IQR)*
Gender, female	14 (46.7%)
Post-test overall score	9 (5.75–10)
Overall score difference (post - pre)	3 (2–7)
Frequency of playing games (“How often do you play entertaining video games each week?”)	
Every day	12 (40%)
3–5 times per week	10 (33.3%)
1–2 times per week	3 (10%)
Not very often	3 (10%)
Not at all	2 (6.7%)

Table 7.6 provides a summary of the descriptive statistics of the IMI subscale, giving the following values for the median and IQR for each statement for each subscale: Pressure/Tension (Median = 1.4, IQR = 1.2–2.55), Interest/Enjoyment (Median = 6.57, IQR = 6.21–6.86), and Value/Usefulness (Median = 6.71, IQR = 6.43–6.75). The low score for Pressure/Tension implies that students experienced low pressure while playing the game, which is a good sign. On the other hand, the other two scales (Interest/Enjoyment and Value/Usefulness) had high values, which is preferable. The reverse statements were reversed to be involved in the aggregative average subscale scores.

Table 7.6: The scores of the IMI subscales.

Statement Number	***scores***
S2	7 (5.75–7)
S6	2 (1–2)
S11	7 (5.75–7)
S13	1 (1–4)
S18	1 (1–2)
Pressure/Tension average	1.4 (1.2–2.55)
S1	7 (6–7)
S4	7 (6–7)
S5	1 (1–2)
S8	1 (1–2)
S10	6 (6–7)
S14	7 (–7)
S16	6 (5.75–7)
Interest/Enjoyment average	6.57 (6.21–6.86)
S3	6.5 (6–7)
S7	7 (6–7)
S9	7 (6.75–7)
S12	7 (6–7)
S17	7 (6–7)
S19	7 (6–7)
S15	6 (6–7)
Value/Usefulness average	6.71 (6.43–6.75)

The scores of the IMI subscales are plotted in Fig. 7.6 and confirm that Interest/Enjoyment and Value/Usefulness are similar, having almost the same median but with a different variation. The same boxplot (Fig. 7.6) shows that Pressure/Tension is different, as should be the case, and has more variation, as shown by the box being longer and the middle line, which represents the median, being lower.

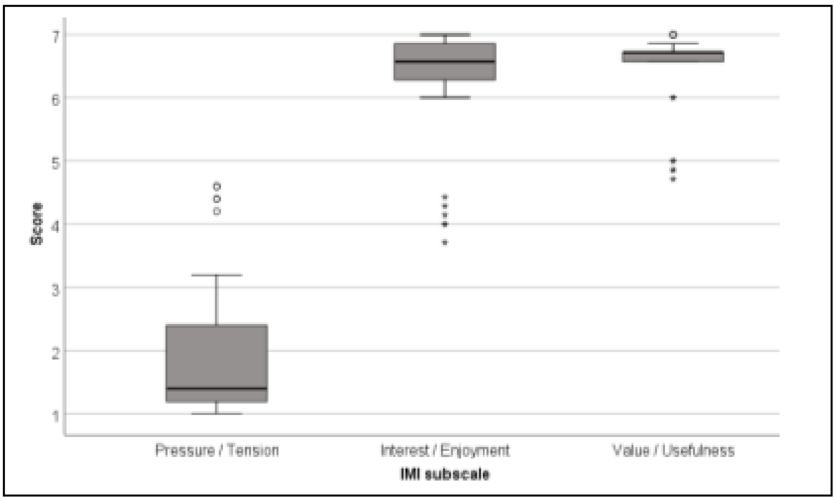


Figure 7.6: The scores of the three IMI subscales.

The output table (Table 7.7) provides Spearman's correlations between all the pairs of IMI subscales, and the results indicate significant correlations where ** indicates significance at the 0.01 level (2-tailed). Table 7.7 shows that the Pressure/Tension score is negatively correlated with both the Interest/Enjoyment (rho = -0.784, p<0.001) and Value/Usefulness subscales (rho = -0.686, p<0.001). On the contrary, the Interest/Enjoyment score is positively correlated with the Value/Usefulness subscale (rho=0.612, p<0.001).

Table 7.7: Correlations between the IMI subscales.

			Motivation	Value	Pressure
Spearman's rho	Motivation	Correlation coefficient	1.000	.612**	-.784**
		Sig. (2-tailed)	.	.000	.000
		N	30	30	30
	Value	Correlation coefficient	.612**	1.000	-.686**
		Sig. (2-tailed)	.000	.	.000
		N	30	30	30
	Pressure	Correlation coefficient	-.784**	-.686**	1.000
		Sig. (2-tailed)	.000	.000	.
		N	30	30	30

Note: **. Correlation is significant at the 0.01 level (2-tailed).

To provide an in-depth understanding of the IMI scores and analyse the outliers in the boxplot (Fig. 7.6), an illustration of each student's IMI subscale scores is displayed in Fig. 7.7. There are two major groups of lines with each line representing a student; the X-axis shows the three different IMI subscales, and the Y-axis shows the scores from

1 to 7. The first group has more lines (students), and both the Interest/Enjoyment and Value/Usefulness scores are close to each other, ranging between 6 and 7.

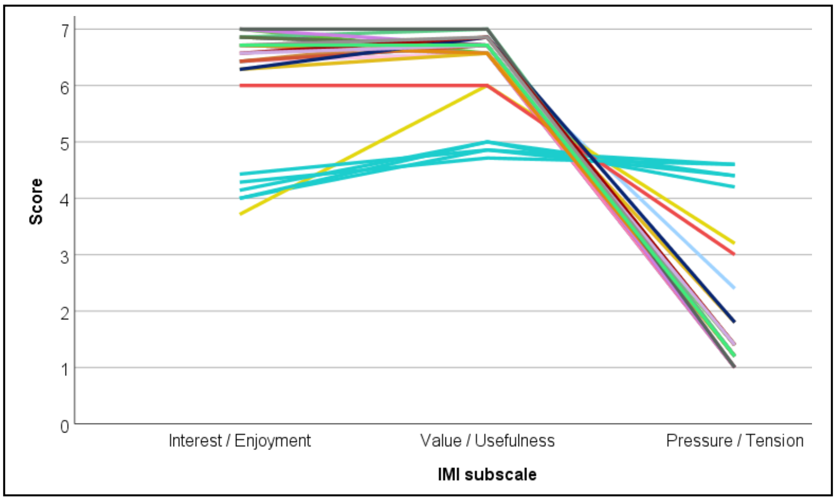


Figure 7.7: IMI subscale scores for each student. (Note: Each line represents one student.)

These scores are much higher than the Pressure/Tension scores, which range from 1 to slightly over 3. The second group of lines is shown in a turquoise colour and consists of the outliers; all of the IMI subscale scores range between 4 and 5. The outlier group could be interpreted as consisting of students who were unable to show their clear opinion about the IMI subscale statements, where a score of 4 means “somewhat true”. Why those students could not disclose their true intrinsic beliefs may be revealed by conducting one-to-one interviews.

The IMI subscale scores will now be analysed in respect to other variables. The first variable is the total score of the post-test, and it was positively correlated with Interest/Enjoyment (rho=0.832, $p<0.001$) and Value/Usefulness (rho=0.682, $p<0.001$) and negatively correlated with Pressure/Tension (rho=-0.946, $p<0.001$). The second variable is the difference score which was calculated as the (post–pre) test scores.

However, there was no correlation between the IMI subscale scores and the total difference of (post–pre) scores ($p>0.05$). These associations are displayed in Fig. 7.8, showing that both Interest/Enjoyment and Value/Usefulness scores increase as the total post-test scores increase. However, the scatter plot in Fig. 7.8 shows that the Pressure/Tension score increases as the total post-test score decreases. On the other hand, Fig. 7.9 shows no patterns between the difference scores and the IMI subscale scores, which means there are no clear associations.

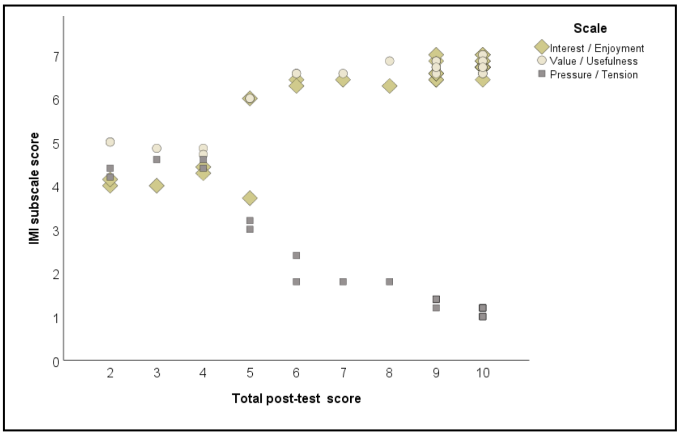


Figure 7.8: The relationship between IMI subscale scores and total post-test scores.

The fourth variable is the gender of learners, and the Mann-Whitney U test was applied to find its relation, if any, to the IMI subscale scores. The result showed no significant association between gender and any IMI subscale score ($p>0.05$ for the three subscale scores). The boxplots in Fig. 7.10 display a comparison of female and male scores regarding each subscale (Interest/Enjoyment, Value/Usefulness, and Pressure/Tension) based on the median values. The median values are almost the same (the line in the middle of each box) in the three IMI subscales, but the male scores show more variability than the female scores.

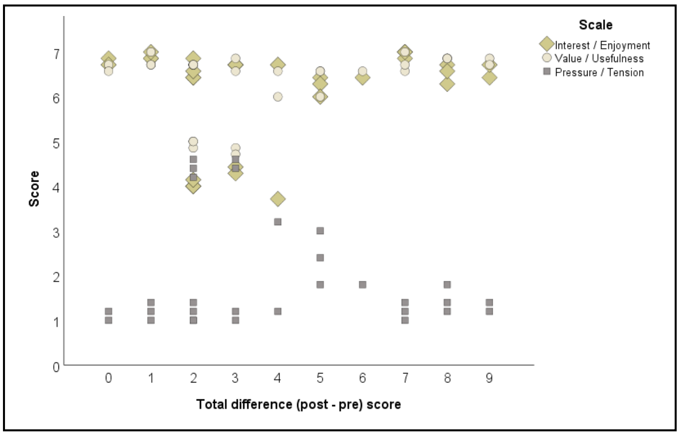


Figure 7.9: The IMI subscale scores in respect to the total difference of (post–pre) scores.

The last variable to analyse is the frequency of playing video games, which has a significant impact on all three IMI subscale scores. It was positively correlated with the Pressure/Tension subscale scores (rho=0.617, $p<0.001$) and negatively correlated with both the Interest/Enjoyment (rho= -0.756, $p<0.001$) and Value/Usefulness (rho= -0.619, $p<0.001$) scores.

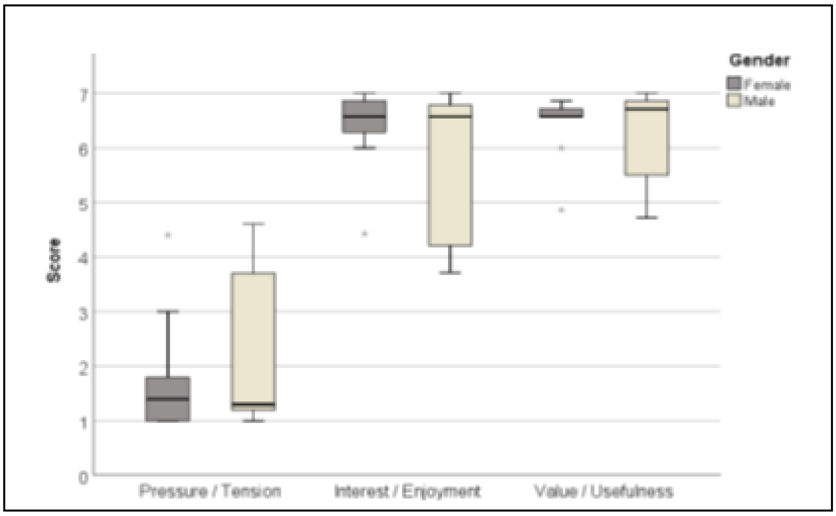


Figure 7.10: IMI subscale scores with respect to gender.

The assumption used while performing Spearman's rho (ρ) was that the distance between successive levels of frequency variable is equal given the following scale: Every day=1, 3–5 times/week=2, 1–2 times/week=3, Not very often=4, Not at all=5. The scatter plot in Fig. 7.11 shows the associations between the frequency of playing video games and IMI subscale scores. The scores of Interest/Enjoyment and Value/Usefulness increase as the frequency of playing video games increases. When the frequency decreases, the Interest/Enjoyment scores show a clear decrease while the effects on the score of Value/Usefulness are limited. Also, Fig. 7.11 shows an opposite association between the Pressure/Tension subscale score and the frequency of playing, where the feeling of pressure scores higher as the frequency of playing video games decreases.

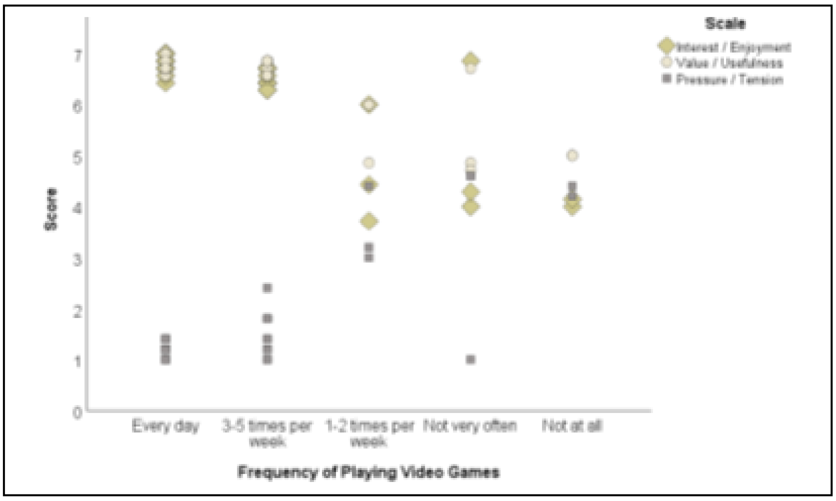


Figure 7.11: The relationship between IMI subscale scores and the frequency of playing video games.

The results of all correlation tests are summarised in Table 7.8 where the columns represent IMI subscales and rows show the variables. In addition, Table 7.8 presents the association between gender and the three IMI subscale scores tested with the non-parametric Mann-Whitney U test, while Spearman's rho (ρ) coefficient was implemented for the remaining association tests. * denotes significance at 0.05 level.

Table 7.8: The associations of the three IMI subscale scores and the variables.

Variables	*IMI subscales*		
	Interest/Enjoyment	*Value/Usefulness*	*Pressure/Tension*
Gender	U=96.5, p=0.515	U=98, p=0.552	U=90.5, p=0.364
Post-test score	rho=0.832, p<0.001*	rho=0.682, p<0.001*	rho=-0.946, p<0.001*
Difference score	rho=-0.077, p=0.686	rho=0.212, p=0.260	rho=0.152, p=0.423
Frequency of playing games	rho=-0.756, p<0.001*	rho=-0.619, p<0.001*	rho=0.617, p<0.001*

Notes. The association between gender and the three IMI subscale scores was tested with the non-parametric Mann-Whitney U test, while Spearman's rho (ρ) coefficient was implemented for the remaining association tests. * denotes significance at 0.05 level.

7.4.4 Findings

The null hypothesis (H_0) of this part of the analysis is that the VFTG, Island of Volcanoes, does not provide motivation for the player. The alternative hypothesis (H_1) is that the VFTG does provide motivation, which is an important element for SGs and GBL. The results of the Interest/Enjoyment scale provide evidence to reject the null hypothesis. The IMI questionnaire was utilised in this research as part of the evaluation process of the proposed conceptual framework. It is a well-established tool applied in a number of studies concerned with measuring motivation regarding activities in virtual environments. Three subscales of the IMI questionnaire were selected: Interest/Enjoyment, Value/Usefulness, and Pressure/Tension, as they were the most suitable for this study and the remainder were ignored either because of the limited time available to the experiment or their unsuitability.

The analysis showed that the majority of learners found the game useful. They were a little indecisive in their opinion regarding motivation compared to usefulness, but it was still considered as providing a great feeling of motivation. Learners showed a low level of pressure/tension, but it was not as minimal as expected, which could be due to a technical issue faced during the experiment. The medians of the scores were as follows: Pressure/Tension = 1.4, Interest/Enjoyment = 6.57, and Value/Usefulness = 6.71. The low median indicates the low level of pressure that learners experienced, which is a positive aspect; on the contrary, the other two medians had high values, which were preferable. Statements S2, S5, S8, and S11 were reversed to be included in the aggregative average subscale scores, and all three subscales displayed high internal consistency and reliability. In summary, the research findings suggest that the game design encouraged intrinsic motivation.

The results revealed that gender differences do not affect the intrinsic motivation of learners. In general, the literature shows clear disagreement over gender differences in educational game motivation, engagement or even learning achievement levels. While this finding is consistent with the results of some studies (S. Kim et al., 2017; O'Reilly, 2014; J. C. Yang & Quadir, 2018), other studies have not considered investigating or addressing this issue (Lieberoth, Pedersen, & Sherson, 2015; Vandercruysse et al., 2013). A simple test of correlation between the frequency of playing video games and IMI subscale scores showed that learners who played every day or three to five times per week had a lower score in Pressure (struggled less) and higher scores in Interest and Value than those who played less frequently (1–2 times per week, Not very often, and Not at all).

7.5 Presence

Presence is defined as the feeling of the player being there in the Virtual Environment (VE) rather than the actual space inhabited by his/her physical body (Grassini & Laumann, 2020; Melo, Sampaio, Barbosa, Raposo, & Bessa, 2016). The sense of presence differs based on different factors related to hardware quality and design elements of the VE (Czub & Piskorz, 2014). There are a number of methods to quantify presence which focuses on objective or subjective measurements (Lombard & Ditton, 1997). Objective measurements include two variants: physiological and behavioural tools. The latter type of objective tools measures the behaviour of the player in VEs while the former type assesses physiological changes in relation to presence such as heart rate, skin temperature, or electroencephalography (EEG) (Hein, Mai, & Hußmann, 2018). On the other hand, subjective measurements, which are the most commonly applied (Grassini & Laumann, 2020), are obtained using questionnaires.

Physiological tools could provide more objective measurements of a feeling of presence that players may experience in the VE. However, the limitations cannot be mitigated, such as the disturbance of players' immersion and the complexity of attaching special equipment to players (Skarbez et al., 2017). Moreover, it would be too costly to make such tools available to school students. There is less literature about behavioural tools and therefore these are rarely applied (Schirm, Tullius, & Habgood, 2019). Behavioural tools are complicated to explain (Schirm et al., 2019) and involve constructing events to be incorporated within the VE with the aim to trigger certain behaviour which could be irrelevant to the VE's main scenario/story (Grassini & Laumann, 2020), and may lead to a reduction in the feeling of presence.

Questionnaires are the most used method for evaluating presence (Hein et al., 2018) and several available questionnaires that measure presence have been proven to be valid and reliable (Skarbez et al., 2017). The Igroup Presence Questionnaire (IPQ) is one of the most frequently used, in addition to the Slater-Usoh-Steed Questionnaire (SUS) (Slater, Usoh, & Steed, 1994) and Presence Questionnaire (PQ) (Witmer & Singer, 1998). Some researchers have used a one-item questionnaire which proved to be reliable regardless of simplicity (Bouchard et al., 2004), but was limited in terms of capturing levels of presence. In fact, presence questionnaires have been built to serve specific applications, such as the Temple Presence Inventory (TPI) (Lombard, Ditton, & Weinstein, 2009), and the ITC-Sense of Presence Inventory (ITC-SOPI) (Lessiter, Freeman, Keogh, & Davidoff, 2001), which focus on non-interactive media while other questionnaires are concerned with social environments (Gerhard, Moore, & Hobbs, 2001; Nowak & Biocca, 2003).

The IPQ was selected because it is suitable for the purpose of this study and applicable to VE and games. However, the PQ (Witmer & Singer, 1998) is the most commonly applied self-report tool to measure presence (Grassini & Laumann, 2020; Hein et al., 2018) but was not selected to use in this research. Looking closely at presence questionnaires reveals variations in the definition of presence among other characteristics such as theoretical foundations, scales, number of questions, and type of Likert scales. The PQ questionnaire highlights the "involvement" and "immersion" traits of the VE, whereas the IPQ is concerned with the feeling of "being there" inside the VE. The IPQ is built on the conceptualization of presence in a VE as a mental model of that VE. A mental model of the VE represents the 3D model and the actions that can be performed in that virtual space. Thus, it is a spatial-functional model while the PQ focuses on involvement and immersion, which are both included in the IPQ (Schubert et al., 2001). In addition, the literature shows that many researchers applied PQ just because of its popular "wide usage" (Hein et al., 2018) without reasonable justification.

7.5.1 igroup Presence Questionnaire

The igroup Presence Questionnaire (IPQ) (IPQ, 1999; Schubert et al., 2001) was applied to evaluate the experience of a sense of presence in the VFTG. The questionnaire consists of a general statement and three factors, which can be viewed as relatively separate factors. These three factors measure the following:

1. Spatial Presence (SP): the feeling of being actually present in the VE.
2. Involvement (INV): the interest in the VE and the participation in the experience.
3. Experienced Realism (REAL): the learner's experience of realism in the VE.

Table 7.9: IPQ statements.

Factors	Statement	Likert Scale
General Item		
G1	In the game world, I had a sense of "being there"	not at all–very much
Spatial Presence		
SP1	Somehow, I felt that the game world surrounded me.	fully disagree–fully agree
SP2	I felt like I was just perceiving pictures.	did not feel–felt present
SP3	I did not feel present in the game space.	fully disagree–fully agree
SP4	I had a sense of acting in the game space rather than operating something from outside.	fully disagree–fully agree
SP5	I felt present in the game space.	extremely aware–moderately aware–not aware at all
Involvement		
INV1	How aware were you of the real-world surroundings while navigating in the game world? (i.e. sounds, room temperature, other people, etc.)	fully disagree–fully agree
INV2	I was not aware of my real environment.	fully disagree–fully agree
INV3	I still paid attention to the real environment	fully disagree–fully agree
INV4	I was completely captivated by the game world.	completely real–not real at all
Experienced Realism		
REAL1	How real did the game world seem to you?	not consistent–moderately consistent–very consistent
REAL2	How much did your experience in the game environment seem consistent with your real-world experience?	about as real as an imagined world–indistinguishable from the real world
REAL3	How real did the game world seem to you?	fully disagree–fully agree
REAL4	The game world seemed more realistic than the real world.	fully disagree–fully agree

The general statement evaluates the overall feeling of being there, and has a great impact on all three factors, particularly on SP. Statements and their factors are displayed in Table 7.9 and all statements have a scale from 0 to 6. To calculate the scores for each factor, first the three negative items should be reversed (SP2, INV3, and REAL1). This can be achieved by multiplying the statement score by -1 and adding 6; the mean score for each factor can then be computed.

7.5.2 Statistical Analysis

The average score for each one of the SP, INV and REAL factors was calculated as the simple average of their corresponding statements. Cronbach's alpha (α) was obtained for each subscale as a measure of internal consistency. Spearman's rho (ρ) coefficient was used to test for pairwise correlations among the three factors and the G1 question. The Mann-Whitney U test was used to examine whether gender is associated with any of the factors, while Spearman's rho (ρ) coefficient was employed to test for correlations between any of the factors' scores and the motivation score, pressure score, post-test score, and the difference score.

7.5.3 Results

The 30 students included in the experimental group filled in the IPQ questionnaire immediately after playing the game. There were no missing values. The IPQ scores were as follows: the G1 question had a median of 5 (IQR = 3.75–6). Questions SP2, INV3, and REAL1 were reversed to derive the aggregative average subscale scores. The SP factor score was 4 (IQR = 3.1–5), while the INV score was 5 (IQR = 4–5.81) and the REAL score was 4 (IQR = 2–4). Fig. 7.12 illustrates the results, with boxplots presenting the scores of the four factors. The figure shows that G1 and INV had the same median (50^{th} percentile) of 5, but the values of INV are not so widely distributed as those of G1, since the box and whiskers of INV are closer to its median and G1 is more spanned. SP had a somewhat lower median (namely, 4) with its values spanned around this central value. The REAL factor had the same median as SP (4) but its values have a downward tendency (skewness to the left) since the box extends to the bottom of the chart.

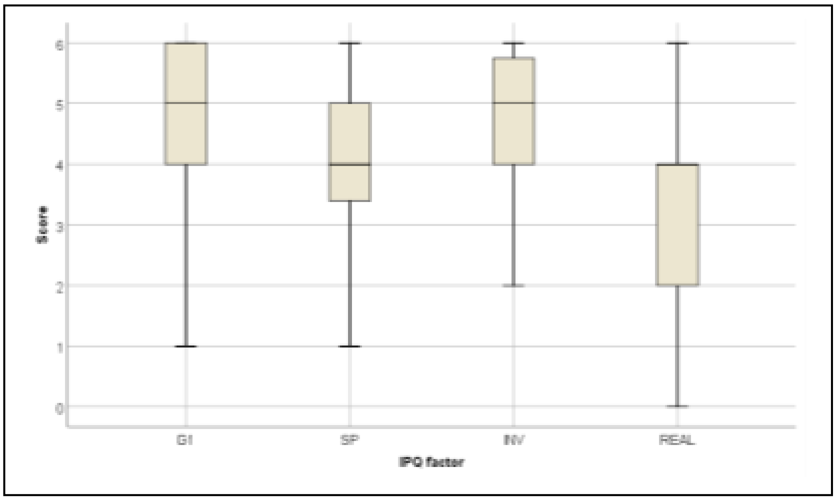


Figure 7.12: Box plots of IPQ factor scores.

The Cronbach's alpha was 0.988, 0.965, and 0.980 for the SP, INV, and REAL factors, respectively, indicating their excellent internal consistency and reliability. The descriptive statistics are summarised in Table 7.10, showing the median and IRQ (Q3–Q1). All pairwise comparisons among the scores (three factors and G1 statement) were statistically significant.

Table 7.10: IPQ scores.

Factors	***Median (IQR)***
G1	5 (3.75–6)
SP1	4 (2.75 –5)
SP2 (reversed)	4 (3–5)
SP3	4 (2.75–5)
SP4	4 (2.75–5)
SP5	4 (3.75–5)
SP	4 (3.1–5)
INV1	5 (4–6)
INV2	5 (4–5.25)
INV3 (reversed)	5 (4–6)
INV4	5 (4–6)
INV	5 (4–5.81)
REAL1 (reversed)	4 (2–4)
REAL2	4 (2–4)
REAL3	4 (2–4)
REAL4	4 (2–4)
REAL	4 (2–4)

The G1 score was positively correlated with SP (rho=0.805, p<0.001), INV (rho=0.859, p<0.001), and REAL (rho=0.793, p<0.001) scores. The SP score also correlated with the INV (rho=0.894, p<0.001) and REAL (rho=0.841, p<0.001) scores. Finally, INV and REAL scores also correlated with one another (rho=0.851, p<0.001). These correlations are presented in the scatter plot matrix shown in Fig. 7.13, where the diagonal shows the distribution of each factor. The medians of the IPQ factors and statements displayed in Table 7.10 show that the VFTG yielded an upper-intermediate level of presence, where all of the IPQ factors (spatial presence, involvement, and realism) have scored over the average. These scores imply that most of the learners considered themselves to have achieved the feeling of being there in the VFTG environment.

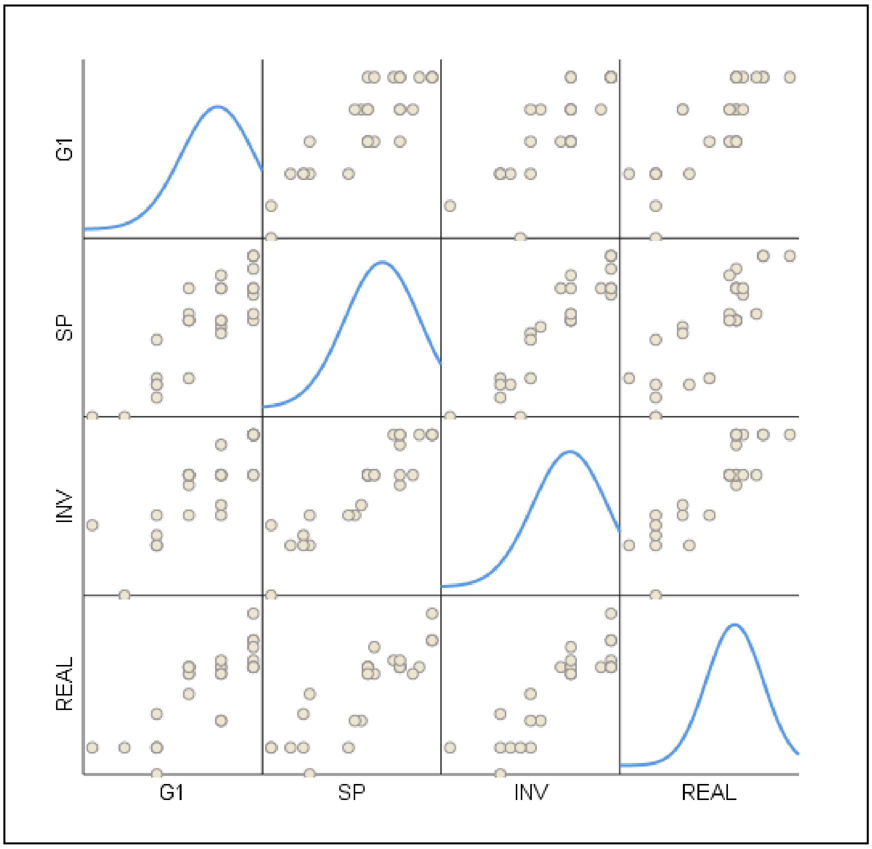


Figure 7.13: Relationships between IPQ factors.

An educated guess can be made that gender difference will not reveal a significant influence on presence in the VFTG as it was not reflected in the motivation data measured by IMI. Gender had no effects on the IPQ score results (p>0.05 for the four scores), as displayed in Fig. 7.14. The learners were split by gender regarding each IPQ factor, including the general statement G1. Both genders had a tendency around 5 for G1, with females achieving a somewhat lower score (around 4.5), but with a narrower dispersion. The difference between genders for the SP, INV and REAL scales was even smaller, as their pairs of medians were at the same level – the same or almost equal medians for the two genders in each factor. In fact, several previous studies have stated that gender difference affects the feeling of being there(Melo, Raposo, Coelho, Narciso, & Bessa, 2019; Melo et al., 2016) and some found that males reported a stronger sense of physical presence (Felnhofer, Kothgassner, Beutl, Hlavacs, & Kryspin-Exner, 2012; Lachlan & Krcmar, 2011; Nicovich, Boller, & Cornwell, 2005).

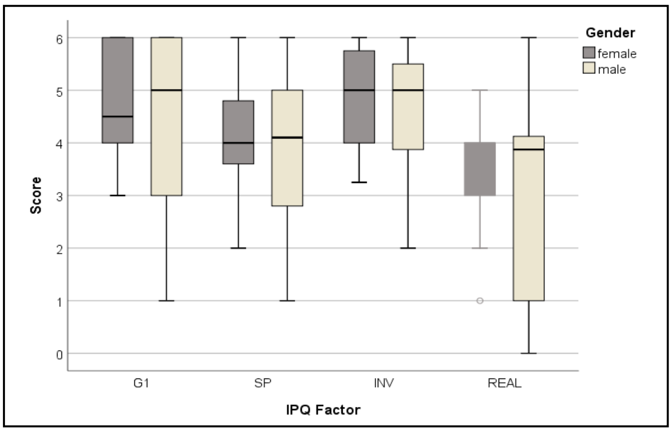


Figure 7.14: IPQ scores with respect to gender.

However, a number of studies agreed with this research result (Felnhofer, Hetterle, Schmidt, Kryspin-Exner, & Kothgassner, 2014; Schuemie, Abel, van der Mast, Krijn, & Emmelkamp, 2005) regarding the role that gender may play in the experience of presence in VEs. Other researchers put forward factors such as exposure time (Melo et al., 2016), stress (Felnhofer et al., 2012), spatial ability (Melo et al., 2019), and previous experience or frequency of playing (Lachlan & Krcmar, 2011) as factors which may play a role in gender differences regarding the experience of presence. Furthermore, potential gender differences have been found in particular with some factors of presence such as realism and involvement in addition to overall presence (Lachlan & Krcmar, 2011). In summary, literature yielded contradictory results regarding gender difference and its effects on levels of presence.

Further analysis could be considered to test the possible influence of the frequency of playing video games and pressure variables on gender differences and their relationship with IPQ scores.

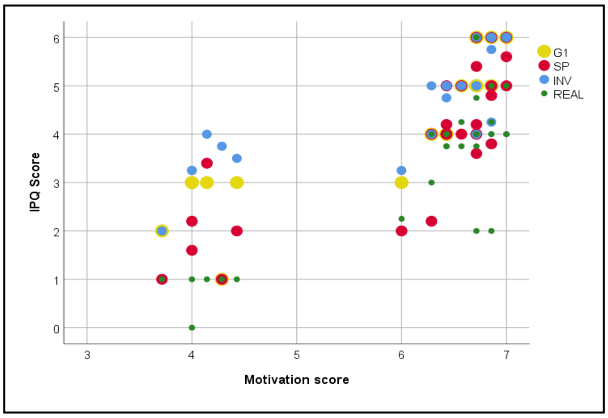


Figure 7.15: Relationship between IPQ scores and motivation scores.

The motivation score, which is the Interest/Enjoyment score from IMI, correlated positively with all IPQ scores: G1 (rho=0.925, $p<0.01$), SP (rho=0.787, $p<0.01$), INV (rho=0.843, $p<0.01$), and REAL (rho=0.692, $p<0.01$). On the other hand, the Pressure/Tension score negatively correlated with all IPQ scores: G1 (rho= -0.776, $p<0.01$), SP (rho= -0.654, $p<0.01$), INV (rho= -0.646, $p<0.01$), and REAL (rho= -0.603, $p<0.01$). These negative relationships are expected given the implication that a low pressure level leads to a higher presence level experienced by learners. Also, the more the learners felt motivated, the more they felt a sense of presence in the VFTG environment.

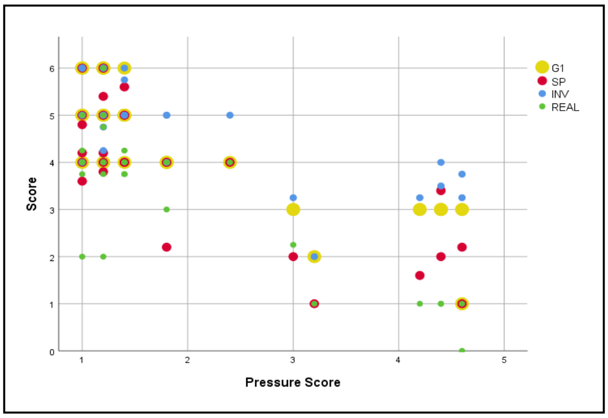


Figure 7.16: Relationship between IPQ scores and pressure scores.

Fig. 7.15 displays the association between the motivation score and IPQ factor scores: as the motivation score increases, the scores of all IPQ factors increase. Fig. 7.16 shows the negative relationship between the pressure score and IPQ factor scores: an increase in the pressure score leads to a decrease in the scores of all IPQ factors. It is worth mentioning that the spacing in data points in both Fig. 7.15 and Fig. 7.16 is explained by the outliers found in the IMI data; there are less than five of these data points. When this cluster of outliers is ignored, the relationships in both figures are more obvious and visible but this cluster does not affect those relationships.

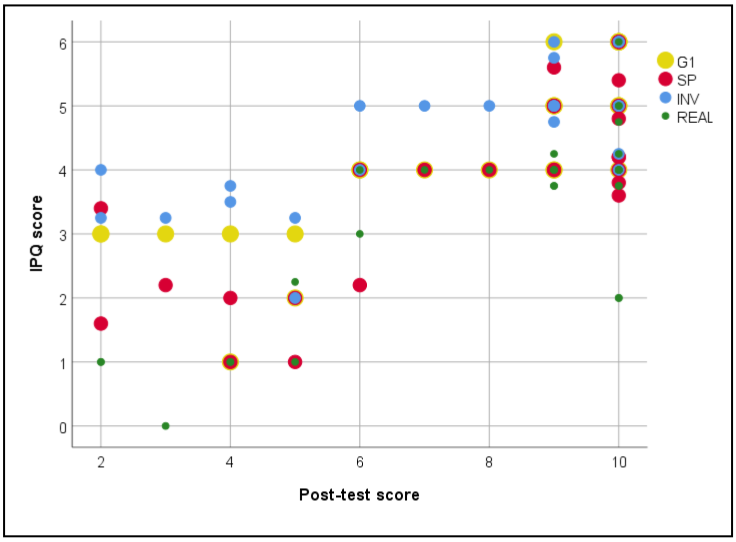


Figure 7.17: Relationship between IPQ scores and total post-test scores.

A similar increasing trend of the four IPQ scores linked to the higher post-test score values is also noted. In particular, the post-test score was positively correlated with G1 (rho=0.852, p<0.001), SP (rho=0.693, p<0.001), INV (rho=0.697, p<0.001) and REAL (rho=0.643, p<0.001). Fig. 7.17 shows the positive relationships between the post-test score and all IPQ factor scores.

On the contrary, the difference scores passed the significance level only for the REAL factor score (rho=0.369, p=0.045), while there was no association with the remaining IPQ factors (p>0.05 for all comparisons). All results are summarized and presented in Table 7.11, showing the relationships between IPQ factors and the study variables.

Table 7.11: Associations of IPQ factors with the study variables.

Study Variables	***IPQ Factors***			
	G1	***SP***	***INV***	***REAL***
Gender	U=106, p=0.797	U=107, p=0.834	U=99, p=0.577	U=102.5, p=0.686
Motivation score	rho=0.925, p<0.01*	rho=0.787, p<0.01*	rho=0.843, p<0.01*	rho=0.692, p<0.01*
Post-test score	rho=0.852, p<0.01*	rho=0.693, p<0.01*	rho=0.697, p<0.01*	rho=0.643, p<0.01*

Notes. Mann-Whitney U test was applied to test the associations between gender and the IPQ score, while Spearman's rho (ρ) coefficient was applied for the remaining correlations. G1, General item representing the "sense of being there". SP: Spatial Presence. INV: Involvement. REAL: Experienced Realism. * denotes significance at 0.05 level.

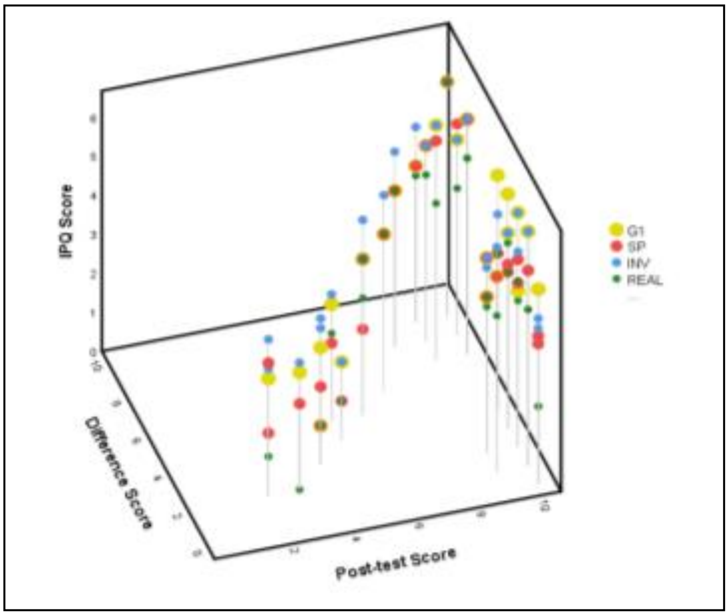


Figure 7.18: 3D scatter plot of independent variables (difference score, post-test score) and IPQ score.

Finding a correlation between only one IPQ factor (REAL) and the difference score raised a question regarding why the other factors did not also present correlations. A further analysis was conducted by plotting the difference score and post-test score as independent variables on the X-axis and Z-axis, respectively, while the IPQ factor score is the dependent variable on the Y-axis.

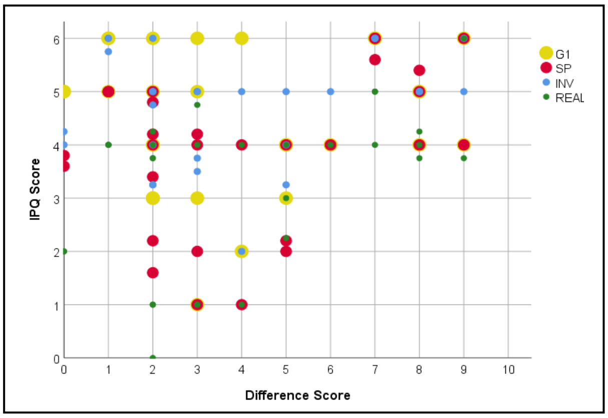


Figure 7.19: Scatter plot of difference score and IPQ factors.

As displayed in Fig. 7.18, there are two data clusters: a small one in the upper outer corner and a large cluster stretching from the upper middle corner to the outer lower middle corner. The shape of this large cluster indicates the possibility of a correlation between the three variables, which means there could be correlations between the IPQ factors and the difference score. Before statistically testing that possibility, the explanation of the small cluster should be clarified. The small cluster represents learners who got high grades in the pre-test (more than 5), and then got higher grades in the post-test. The data of those learners, which form the small cluster, could be excluded while testing the relationship between the difference score and IPQ factor scores.

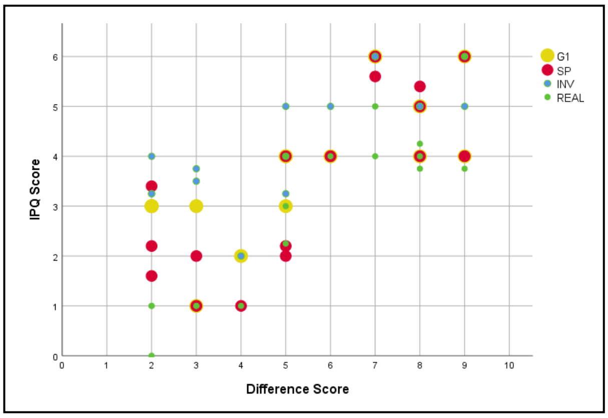


Figure 7.20: Relationship between IPQ scores and difference scores.

The overview of the whole dataset displayed in Fig. 7.19 reveals no association between the difference score and IPQ factor scores with the exception of a weak correlation with REAL score, as already proved statistically. Fig. 7.20 illustrates clear positive correlations between the difference score and all IPQ factor scores after eliminating the small cluster. The difference score is now positively correlated with G1 (rho=0.765, p<0.01), SP (rho=0.747, p<0.01), INV (rho=0.750, p<0.01), and REAL (rho=0.807, p<0.01). All IPQ factors have a strong correlation with the difference score, especially the REAL score. Looking back to the small cluster, the medians for the IPQ factors are the same as the whole dataset, revealing some outliers associated with the REAL score.

7.5.4 Findings

The null hypothesis (H_0) of this part of the analysis is that the VFTG, Island of Volcanoes, does not provide a sense of presence for the player. The alternative hypothesis (H_1) is that the VFTG does provide a sense of presence, which is an important element for SGs and FBL learning. The results of the IPQ factors provided evidence to reject the null hypothesis. The IPQ tool was applied to measure the feeling of presence in the VFTG, Island of Volcanoes, as part of the evaluation process of the proposed conceptual framework utilised to design and develop the VFTG. Regarding the general feeling of presence, learners mostly manifested a sense of “being there” in the VFTG environment, as verified by their responses to G1 with a median equal to

5, which was above the average and considered a high score along with the INV factor. Thus, there is a clear indication of presence in the Island of Volcanoes as a feeling of being there. To a slightly less significant extent, the SP and REAL factors also presented good scores where the medians were above 3 on a 0 to 6-point Likert scale.

The level of perception regarding SP factor indicated that most of the learners experienced physical presence in the virtual space and felt a sense of action. The overall score of the SP factor equalled 4, which was above the average, that is disclosing a moderate extent of simulated authenticity of the actual FT scenario in Island of Volcanoes, the VFTG. The INV score suggests that learners were interested in the VFTG environment and showed how involved they were in the virtual trip experience instead of being distracted by the physical world around them; the median equalled 5. The learners experienced a proper level of realism inside the VFTG where the REAL score was higher than the average but with a wider variance than the rest of IPQ factors, which made it less significant.

The gender difference indicated no effects on the feeling of presence in the VFTG, while further analysis could clarify the influence of other variables such as frequency of playing video games and pressure on gender difference and the impacts on IPQ scores. On the other hand, there was a clear relationship between the post-test score and IPQ factor scores where learners who got high scores in the post-test revealed a stronger feeling of being there than those with a lower post-test score. The difference score, calculated by subtracting the post-test score from the pre-test score, showed an association with the IPQ scores after analysing the whole dataset into two clusters. For the cluster of learners who achieved high scores in the pre-test and got the same high scores or even higher in the post-test, their difference scores were eliminated. The reason for this elimination is that their difference scores were low while their IPQ scores were high while the stronger pattern connects the large difference with high IPQ scores and the low difference score with low IPQ scores.

7.6 In-game Assessment

This section explains the plan to analyse and assess in-game data for future work. The in-game data is collected based on the interactions and indicators that are defined by the evaluation model. The process of collection is already explained in the previous chapter (see Chapter 6, Section 6.4). The conceptual framework provides the foundation of in-game assessment for VFTGs via the evaluation model, and game designers/educators can adapt the analysing process to their learning goals and research questions. As a part of this study experiment, the in-game data is logged from the experiment group (30 students). The first step is mapping the post-test questions to the knowledge concepts (Ks) that are utilised in designing the learning tasks (LTs) of the VFTG prototype and presented in Table 7.12 to facilitate the comparison between in-game and after (post-test) assessments. The K_5 excluded as it is about the general knowledge of playing video games. Some knowledge concepts are utilised by more than one question of the post-test as Q3 and Q1 require the same knowledge K_4 and Q9 and Q10 require K_6 .

Table 7.12: Mapping post-test questions to *Ks.*

Post-test Question	Knowledge
Q2	K_1
Q10	K_2
Q6	K_3
Q3 & Q1	K_4
Q9 & Q10	K_6
Q8	K_7
Q8	K_8
Q5	K_9
Q4	K_{10}
Q7	K_{11}

The knowledge indicators are defined for three practice tasks (LT_1, LT_2, and LT_3) and three assessment tasks (LT_4, LT_6, and LT_7, $SK(LT_8, LT_9, LT_{10},$ and $LT_{11})$). The in-game data of two practice tasks (LT_1 and LT_3) were collected while it was difficult to collect any data regarding LT_2 (see Chapter 6, Section 6.3.2.3). Designing LTs according to interactions and indicators gives educators the flexibility to assess tasks whenever the required indicators' data is collected. For that, LT_1 and LT_3 can be assessed.

The FB indicators are considered to find if the player made interactions that reflect his/her FB knowledge. The collected in-game data captured the two types of FB indicators: observational and spatial indicators. The observational indicators are defined for two LTs (LT_5 and LT_8), while the spatial indicator is captured for the whole learning experience. The time indicator is considered for three LTs (LT_6, LT_7, and SK), and for the whole learning experience from the start of play to the end. The decision-making skill indicator is specified for SK which is composited of four LTs each of which (LT_8, LT_9, LT_{10}, and LT_{11}) represents one step of performing the skill.

LT_1 has two Ns, but only N_2 can be assessed while N_1 won't as it requires search action and performing N_2 via collecting action reflects performing N_1. The spatial tracking could be utilised for assessing the search action, but the survival kit items are spawning randomly on the island. The assessment method $S(\,)$ of N_2 can specify some criteria such as collecting a particular number of items and scoring the player differently according to the number of collected items.

The assessment of LT_3 is similar as the task has two Ns; N_1 can be assessed where the data of its action is collected which is flagging the action of clicking the link to watch a YouTube video while N_2 can't be assessed as there is no data collected about the actual watching action.

For both LTs, the weights should be recalculated to give 1 to the assessed Ns. The evaluation model provides flexibility to game designers/educators to define $S_2(\,)$ to be suitable to their learning objectives.

$$LT_1 \;=\; [0.5, 0.5\,]\;[N_1, N_2]^T$$
$$AR \;=\; [0{,}1][S_1(N_1), S_2(N_2)]^T$$

The player has to collect at least two items and can increase the score by collecting more.

$$LT_3 \;=\; [1{,}0]\;[N_1, N_2]^T$$
$$AR \;=\; [1{,}0][S_1(N_1), S_2(N_2)]^T$$

$S_1(\,)$ of LT_3 assesses the player as simple as performing the interaction or not.

LT_4 consists of two interactions which both can be assessed based on a specific criterion such as the counts of collecting gems and putting out fires. In addition, the player should score 10 or more in putting out fires to access the following task. LT_5 is assessed based on correct or incorrect answer and can be analysed regarding to the movement path (spatial indicator) of the player to find out if there is any correlation between the player's movement, which reflects exploration, and answering correctly as the answer depends on exploring and observing.

The time indicator for LT_6 and LT_7 is collected as the completion time (CT_{LT_6} and CT_{LT_7}) of all Ns that performed without distinguishing the time of each N individually. The time indicator of LT_7 can be utilised in assessing the player's performance based on the time spent in searching and collecting the required information which is more logical than assessing the performance of LT_6 in relation to time as solving the task depends on recalling previous knowledge. The same concept of assessment can be applied to SK regarding the time indicator as making a suitable decision in a shorter time should be rewarded.

After computing AR for all LTs except LT_2, the total result of assessment can be compared to the total of post-test to prove that the in-game assessment is effective and can replace the after FT assessment. In addition, the movement path of players, the duration time of play from login to logout, and the proficiency levels (PLS) can be utilised to further assess the player performance behind the capability of after FT assessment.

7.7 Summary

- The evaluation of the proposed conceptual framework included two methods: the experts' perceptions of its connection, usefulness, and usability (see Chapter 5), and the effect of the VFTG prototype that designed and implemented based on the framework as displayed in this chapter.
- The impact of the VFTG prototype reflects on the validation of building linkages between FBL and game design aspects implicitly and the effectiveness of designing and implementing of VFTGs based on the proposed framework explicitly.
- The impact of the VFTG prototype is measured by the learning performance and the learning experience (motivation, and presence).
- The selection of measuring the learning performance is expected as the main goal of providing VFTG is enhancing learning performance.
- The selection of motivation and presence is based on their positive effects on students' performance.
- Considering three variables to evaluate the conceptual framework lies on the importance of their combination on the VFTG experience.
- Qualitative data such as interviewing or observing the students could provide valuable insights about the impact of the VFTG prototype and therefore about the validation of the proposed framework. However, collecting qualitative data means that students should skip their classes to take part in one-to-one interviews or observation sessions, and it would lead to placing demands on the time available for conducting the experiment which is limited based on the time schedule of teaching. However, there was an attempt of running a one-

group experiment involving undergraduate students from the university where the participants were observed and interviewed after playing the VFTG but it was interrupted by the first lockdown and the size sample was not enough to draw any conclusion.

- It is worth mentioning that the time interval between the pre-test and post-test was carried out four weeks after the educational intervention which does not reflect the best practice. The post-test should be completed by students immediately after the intervention and a retention test could be performed after four weeks.

The fourth and final question (***To what extent can the proposed conceptual framework improve the learning process of virtual field trip games?***) of this thesis is answered via the main findings of evaluating the conceptual framework based on the three variables as follows:

- Learning performance:
  - A similarity between participants is established showing control over some variables (ability, age, and percentage of females and males) to ensure that they would not impact the learning performance.
  - The design of the VFTG prototype improved learning performance to a greater extent than did the traditional method of learning, which is demonstrated by the LLTM results, revealing the positive effect of the proposed conceptual framework.
  - The “Understand & Apply” skills along with the “Evaluate & Analyse” skills are enhanced compared with the traditional method of learning. However, the “Remember” skill stayed the same in the VFTG prototype and traditional learning.
- Motivation:
  - The VFTG design encouraged intrinsic motivation as the majority of learners found the Island of Volcanoes useful and provided a great feeling of motivation and low level of pressure as disclosed by the median scores of the three selected subscales (Interest/Enjoyment, Pressure/Tension, and Value/Usefulness) of the IMI questionnaire.
  - A desirable effect of the VFTG design on learning performance is concluded from the positive correlation between the post-test score and Interest/Enjoyment and Value/Usefulness scores, and the negative correlation with Pressure/Tension.

  - The frequency of playing video games reflects on the IMI subscale scores where the players who played every day or three to five times per week felt less pressure (struggled less) and more Interested and Valued the VFTG experience than those who played less frequently.
  - The results revealed that gender differences do not affect the intrinsic motivation.
- Presence:
  - The VFTG design promoted the feeling of presence as learners mostly manifested a sense of "being there" in the VFTG environment which is implied by over the average scores of all IPQ factors (spatial presence, involvement, and realism).
  - Learners felt a moderate extent of simulated authenticity of the actual FT scenario in the VFTG, which is disclosed by the score of SP factor.
  - Learners were interested in the VFTG environment and shows how involved they were in the virtual trip experience instead of being distracted by the physical world around them which is suggested by the INV score.
  - The learners experienced a proper level of realism inside the VFTG which is reflected by REAL score.
  - The gender difference indicated no effects on the feeling of presence in the VFTG.

    learners who got high scores in the post-test revealed a stronger feeling of being there than those with a lower post-test score.
- A plan for analysing the log data, which is collected based on the defined interactions and identified indicators, is explained for future work. The plan shows the possibilities of utilising the log data for an alternative assessment to after the field trip assessment.
- The VFTG provides an opportunity for living field trip experience when a physical field trip is impossible to deliver. The main assumption identified at the beginning of this research is that the VFTG method is not a replacement method for physical field trips but is an alternative method.

The results of the two studies (preliminary and main) support the contributions of this thesis. The experts' perceptions, the improvement of learning performance, and signs of learners' motivation and feeling of presence confirm the linkage between FBL and game design aspects via the three links (ELT modelling, Game modelling, and World modelling). The enhancement of "Understand & Apply" skills along with the "Evaluate & Analyse" skills shows that the task modelling, evaluation model, and skill modelling are effective in teaching the required skills of FBL. These models are built on mapping the ELT to the internal economy mechanic (ELT modelling) which indicates the provision of the learning experience. The great feeling of motivation and the sense of “being there” points out the efficiency of the matching scheme tool and the world modelling. Finally, the collected log data and the proposed plan of analysing it shows partially the value of standardisation and generalisation of assessment via indicators that can be captured from interactions forming learning activities. The indicators (knowledge, FB, and skill) are provided by the evaluation model to assess the learning tasks of FBL. These indicators facilitate the design and implementation of in-game assessment regardless of the subject or players of the VFTG.

8 Conclusion and Future Work

8.1 Introduction

This thesis focused on developing a conceptual framework for designing virtual field trips in game environments to assist students' experiential learning. To structure a conceptual framework that supports game designers/educators, the learning aspects and game design aspects related to field-based learning (FBL) were defined and three links were built to connect these aspects with the aim of designing virtual field trip games (VFTGs). The first link maps experiential learning theory (ELT) to the internal economy mechanic via game machinations, while the second link transfers the learning process into gameplay through a matching scheme tool, task model, evaluation model, and skill model. The matching scheme tool defines the structure of VFTGs by adapting FBL structure to fit the virtual settings. The task model helps to shape and link tasks to suitable game mechanics in relation to FBL, while the evaluation model develops the assessment process for the designed tasks based on the internal economy mechanic and specific resources (indicators). The third link defines the variables of world modelling to connect the field environment to the game environment. In general, designing VFTGs does not aim to replace the physical field trip (FT), but to provide an effective alternative solution when it is difficult to conduct a physical FT.

8.2 Research Contributions

Designing VFTGs requires the guidance of frameworks to ensure the effectiveness of the designed learning experience. Some of the existing frameworks target the design of game-based learning (GBL) without consideration of FBL pedagogical aspects, while other frameworks which focus on designing VFTs consider limited game design and learning aspects and in the main do not connect these aspects. The importance of VFTGs comes from providing more than a high-fidelity presentation of a physical FT and this thesis proposes a conceptual framework which facilitates and enhances the quality of GBL design for VFTs. Utilising game mechanics and elements in designing VFTs will engage and motivate learners to interact with the VFTG environment and perform virtual fieldwork.

The proposed conceptual framework consists of three conceptual links that connect the learning aspects of FBL to game design aspects; these constitute the contributions of this thesis:

ELT Modelling

ELT creates a cyclic path of learning: observing an experience, reflecting on it, analysing the data and synthesizing, and then testing by doing. The learning process is as important as the outcome of learning; the direct experience and ability to interact directly with learning content in relevant contexts are achievable in FTs. ELT modelling supports the design of the cyclic path, the task and evaluation models define and capture both the learning outcome and process, while direct interaction is modelled through all three links.

ELT modelling maps ELT into the internal economy mechanic, which is employed to quantify ELT concepts in order to connect each stage to the game design. Three components (resources, internal mechanics, and feedback loop) transfer the theory to the game's internal economy. In essence, ELT aims to develop the progression of learning via experience and reflection on doing. Resources are defined, along with a suitable internal mechanic and feedback loop when needed for each stage of ELT. As a result, building blocks of experiential learning are formed based on modelling each stage of ELT. The building blocks facilitate the process of designing VFTGs.

While performing learning tasks, ELT modelling shows the player's progress in reaching different stages of the ELT cycle by producing resources and displaying their flows. Four resources are identified, each of which symbolizes a different level of achievement and is related to a particular stage of the ELT cycle. The levels of achievement are ***level1***: observing, ***level2***: reflecting, ***level3***: synthesizing, and ***level4***: doing. Internal mechanics (source and converter) are employed to display the flow and transaction of these resources from one stage to the next in the cycle. Solving a learning task and constructing new knowledge in steps by going through the cycle can be guided via feedback to improve performance in the next cycle by acting on the learning feedback.

ELT modelling starts with concrete experience (CE) because this is the most frequent entrance stage of the ELT cycle. However, the ELT modelling characterizes the building blocks of each stage, meaning that the game designer/educator can reorder the building blocks to start the ELT cycle from any stage based on the player's learning style.

Game Modelling

Game modelling helps game designers/educators to connect the learning process to gameplay through four steps: matching scheme tool, task model, evaluation model, and skill model. The first step involves employing the matching scheme tool to shape the VFTG's structure by adapting the FBL structure to the virtual environment context and then linking it to suitable game elements. The second step shifts the designing process to focus on learning tasks via connecting learning elements (goal, objectives, outcomes, assessment methods) to game mechanics and selected game elements from the matching scheme. The third step develops the assessment via the evaluation model to benefit from the characteristics of VLEs. The fourth step focuses on specific types of learning tasks which involve high-order skills such as decision-making and teamwork skills.

<u>Matching Scheme</u>

The matching scheme tool was developed in the form of a table to support game designers/educators to transform the FBL structure into a virtual field structure in the game environment. It supports defining the VFTG's structure and connects it to the required game elements. The game elements are selected based on their ability to enhance the learning experience of VFTGs and the relationships between them. The tool is associated with three phases established based on the FBL structure with adaptation to fit the VFTG environment: pre-, during, and post. Each phase is split into a number of steps, each of which is linked to a number of game elements to assist building the FT experience in VFTGs. It is advised to not skip any of the steps, but game elements of each step can be selected to match the learning objectives, and the order of steps in each phase is not restricted. Utilising the matching scheme will advance the development and connection of FBL to the game design aspects.

The pre-phase consists of four steps: start, pre-task, complexity of task, and skill. The start step focuses on ways to adapt the VFTG to each individual player by collecting information in the player's profile, specifically on learning style and prior knowledge. The character of the player should also be considered to allow players to define their identity, giving them control of choice and connection to the narrative. The pre-task step involves preparing the player to perform the task through training on a similar task or showing an example of a solution and can be accomplished by delivering a tutorial level or instructions. Also, a gripping narrative or theme can be introduced from the start of the VFTG to establish a feeling of immersion and engagement. The multimodal presentation should support all of these elements (tutorial level, instructions, or narrative). Task complexity is an essential preliminary step for the following phases and can be controlled by two variables: location and time for VFTGs. Time is a variable of any learning task but combining time with location is distinctive to FBL. The location is a constant variable in traditional learning (classroom) but in online learning, it can be anywhere. However, the variability of location does not add value to the learning task or outcome. On the contrary, FT activities are developed in different locations which deliver the required learning content and context. Identifying higher-order skills is the last step in the pre-phase, such as decision-making and the supported game elements (time, control of choice, resources).

The during phase focuses on defining the learning task based on the required game elements, as it is important for both learning and gameplay. The learning task is split into subtasks representing ELT stages (CE, RO, AC, and AE). These subtasks should provide a progression mechanism to enable the player to go through the entire ELT cycle. In the second step after employing ELT modelling to define each stage, the game designer/educator needs to link each individual stage to the appropriate game elements by facilitating the matching scheme tool.

The CE stage needs interaction, mystery, multimodal presentation, narrative, choices, and emotions. Interaction supports discovery while including some mystery would prompt curiosity and more engagement to explore the VFT environment. Multimodal presentation is important in designing subtasks of all stages; however, in the CE stage it ensures the motivation to explore while interacting with the VFTG environment to observe and collect data. Narrative produces an engaging and immersive context and acts as a hook to complete the following stages. Also, the narrative helps to create more choices and the player’s choices might change the narrative, which would result in creating more choices; it is a positive feedback loop. Adding emotions to the learning experience improves the ability to recall knowledge/skills and strengthens the connection to the learning experience.

The RO stage requires control of choice, challenge, and interaction. The player should have control of choice to reflect on the previous subtask, which varies from one player to another. The challenge motivates the player to reflect, which is essential to ensure progression to the next stage. A balance between choice and the necessity of reflection is required. Interaction in RO supports finding more information to reflect or to ensure the accuracy of reflection by re-interacting with CE events.

The AC stage requires interaction, challenge, and resources. Interaction supports the process of synthesizing concepts. Synthesizing can be accomplished by challenging the player to abstract a concept experienced in previous stages, for example by communicating with a peer in a multiplayer VFTG, chatting with an NPC, or formulating a model in a notebook. The player could need more resources (observed information from the VFTG environment – the colour of plants).

The AE stage needs a challenge, control of choice, interaction, uncertainty, and chances. This stage challenges the player to employ what has been experienced from the previous stages. Consequently, offering control of choice to complete the subtask is derived from the player's understanding of the three previous stages. Since this stage concentrates on doing, all types of interactions should be utilised to test the synthesized concepts/theory. The uncertainty of what comes next in the AE stage drives the learning experience with the element of chances to analyse and question each step of solving the subtask before doing it. Positive or negative consequences can be utilised to guide the player through the experience instead of waiting for the feedback of the completed cycle.

The post-phase concentrates on assessment and its associated concepts (feedback and progression). It typically takes place days or weeks after completing the actual FT, which could impact the assessment and decrease the advantages of feedback. This framework considers assessment as an integrated part of VFTGs to guarantee the usefulness of feedback and adaptation. Formative assessment plays an important role in learning as the feedback improves learning in continuous ELT cycles by bridging the gap between actual and desired performance. Assessment and feedback can be given for each stage or the complete ELT cycle. Experience points (XP) or any resources that can be earned but cannot be exchanged are suitable to reflect grades of evaluation. Recognition by game elements such as trophies, badges, skins, or calling cards promotes social interaction and competition. Evaluation should be associated with feedback and a broad variety of game elements can be applied to deliver feedback with multimodal presentation (audio, text, image, video). In VFTGs, feedback can be embedded into the VFTG environment to catch the player's attention and motivation, such as changing the colour of an object (animal, or artefact in a museum) or stimulating a breeze in a particular direction as a hint. Also, feedback can be provided in the form of a report after completing the ELT cycle. However, for the feedback to be effective the player has to act on it. Yet, the player should be given the choice to take action.

The progression of the player influences their encouragement and motivation to continue with the learning process. Game elements can be utilised to show progression to the player, such as a progression bar and unlocking content. The progression bar works as an acknowledgement of the player's progress in the VFTG; the player can be driven by good progress or pushed to work harder if they have made slow progress. Unlocking content (new tasks, tools, or learning materials) shows progression and motivates the player to unlock additional content.

Task Modelling

Tasks in VFTGs should be formed based on the understanding of all the stages of ELT and designed to allow the player to go through all of the stages. Task-based learning (TBL) is a learning theory that employs the task as a unit of analysis where the syllabus/instruction is broken down into units (tasks) to organize learning; the focus is on the process rather than the final product. Two main concepts are adopted from TBL: 1) the task forms the unit of designing learning in VFTGs and 2) there is a focus on the learning process along with the final product. This conceptual framework considers the task as the unit of learning by arranging the design of VFTGs around tasks and models the design of tasks to capture the process as well as the final product. Therefore, a task should be designed to allow the player to analyse a series of interactions based on the required knowledge in order to solve the task. This results in establishing an overlapping network of interactions to accomplish the desired outcomes of the learning process. The task should be designed with control of choice to solve the task in different ways (different sequences of interactions). The task model consists of three steps: task preparing, task design, and task evaluation.

Task preparing involves defining some learning elements (goal, objectives, complexity, and difficulty) that are essential for designing VFTG tasks. The learning goal and objectives have to be specified as a first step. Complexity differs by manipulating time and location: less complex tasks are accomplished in a short time and one location but the complexity can be increased with a longer period of time and more locations. For instance, extended locations and long periods of time permit the design of tasks for more than one ELT cycle. A simple diagram is displayed in Chapter 4 to guide the game designer/educator to select the right combination of time and location values (short/long and single/multiple) for the learning goal and shows the expected features of the resulting VFTG (number of tasks, type of tasks, single/multiple players, type of narrative, type of guiding). The difficulty of the task should be defined and increased in accordance with the player's progress and level of ability.

The task design step includes identifying the game mechanics and learning outcomes and connecting them to selected game elements based on the matching scheme. The game mechanics (running, searching, role-playing, puzzle solving, constructing, social interaction, and tactical strategy) should be selected to support the learning elements that were defined in the first step, such as applying a searching game mechanic for an observing learning objective. Then the game mechanic can be connected to appropriate game elements from the matching scheme. For example, multimodal presentation and interaction can enable a search mechanic. The design of a learning task in the VFTG requires learning elements, game mechanics, game elements, and learning outcomes. The learning outcome is fed into the task evaluation step. Some examples are given (see Chapter 4, Table 4.2), which can be consulted to design two types of learning tasks in relation to ELT: grasping experience (CE & AC) and transforming experience (RO & AE). Educators can define the learning outcomes and difficulty levels to match the subject of the VFTG. However, the revised Bloom's Taxonomy and its levels of difficulty are applied to the examples given. A story task is an engaging way to connect learning tasks, such as the threat posed by some insects, where for every plant collected by the player, it increases the rate of producing this threat.

Task evaluation is the final step in task modelling, and in this step a decision about the assessment method should be made based on the results of the previous two steps of task modelling. Evaluation can be implemented after completing the VFTG or as an in-game assessment, which can be of two types: embedded (capturing interactions) or traditional tasks (multiple choice). Also, learning tasks can be designed for practice purposes without grading or for assessing performance and defining the learning gap. Assessment after completing the VFTG/physical FT usually ignores evaluating the learning process and concentrates only on the learning product. It is challenging to assess the learning process in physical FTs because of the increased number of learners and the limited time available. However, in-game assessment can overcome these limitations by collecting data of both the learning process and the final product. Regardless of the in-game assessment method (multiple-choice questions or measuring a game object), the assessment task should be matched with game mechanics/elements such as the narrative which enhance the possibility of interweaving the task into the VFTG environment with the aim of reducing learning interruption. The details of the in-game assessment are designed by the evaluation model. Therefore, in the case of applying external assessment after completing the VFTG, the evaluation model can be skipped. However, if the assessment has already been written, the evaluation model can be utilised to convert the test into an in-game assessment.

Evaluation Model

The evaluation model helps game designers/educators by two means:

- ✓ Understanding the evaluation process as an internal economy mechanic by defining resources, internal mechanics, and feedback loops. The internal economy mechanic is explained in steps forming building blocks that can be employed by game designers/educators. The identified resources show the main components of evaluation which would appear in most evaluation processes. In addition, the defined feedback loops display the mechanism of balancing whenever needed, such as balancing acting on feedback and the achievement/scoring of the next task, which would be expected to increase, while not acting on the feedback would lead to decreasing the achievement/scoring of the next task. The internal mechanics such as solving, assigning, and generating feedback should be replaced with suitable internal mechanics for the designed VFTG. For example, the search mechanic can be used instead of the *Solve* converter, and a specific assessment method used instead of *Assign* source.
- ✓ One of the resources produced from the evaluation process based on the proposed evaluation model is indicators. Learning tasks can be very distinctive to a specific game or study subject. However, defining learning tasks as a series of interactions expected to be performed by the player leads to generalisation and standardisation of indicators. The standardised indicators facilitate the process of assessment regardless of the FT subject. The indicator resources are defined to distinguish three types (knowledge, FB, and skill). These three types of indicators support game designers/educators to identify the appropriate evidence to be found as a sign of constructing knowledge/skill. These indicators are generic and can be utilised to design the evaluation of any VFTGs and also are explained with the required context to give game designers/educators a clear guide to design in a process extending from learning goals to results.

Skill Model

The skill model explains a way to design learning tasks that target higher-order skills by providing two examples of skill modelling. These two skills (decision-making and teamworking) are important for FBL and both models can be adapted to include other skills. The skill model is built on the concept of the internal economy mechanic by modelling the unique features and requirements of each skill individually and both skills can be designed in the same VFTG. For example, designing the VFTGs to provide a direct experience in a social setting increases the effectiveness of learning as knowledge is constructed in the company of others and through observation of others' experiences. Social interaction inside a VFTG, along with gaining teamwork skills, could lead to players' appreciation of a community of practice. Social interaction is involved in the matching scheme tool and the evaluation model, while the social setting was modelled as a part of the teamwork skill model. Each skill is modelled in building blocks to support flexible design as more blocks can be added.

World Modelling

The world modelling connects the field environment to the game environment by defining variables and rules to form the VFTG world. The key variables are space, time, and storytelling, and the rules establish relationships between the main variables themselves and variables from the matching scheme tool and evaluation model. Based on the rules, the concept of designing VFTGs in relation to the internal economy mechanic will blend in with the VFTG world. For example, the space and time variables are determined by Ks which are identified as a part of the evaluation model, while storytelling evokes emotion which should be identified by the matching scheme tool. Also, the space and time variables produce spatial and temporal tracks which can be employed by the evaluation model.

The world modelling variables and the connections to learning tasks via the process of modelling ELT, task, evaluation, and skills form the scene/level of VFTG as combined in the final equation (Eq. 4.23):

$$S = gt + \sum_{i=1}^{n} O_i + \sum_{i=1}^{r} LT_i \ , \ gt \in TI$$

Where gt is the unit of time, O is the unit of space, and LT is the learning task.

8.3 Limitations and Future Work

Although the experiment of the thesis was conducted effectively, some limitations should be summarized:

- The number of participants in this research is reasonable in comparison to previous studies even though the population sample is small. However, the sample size could be considered as a limitation. The sample of the preliminary study included 23 participants of experts (game designers and educators) and the main study recruited 60 secondary school students. Another consideration involving the population is that it could not be extended to students from different cities or countries due to time and resource barriers.

- Another limitation is the time gap between the educational intervention and completing the post-test. Immediate performance of the post-test may further ensure that extraneous variables will not have influenced the results. The best practice is to complete the post-test by participants immediately after the intervention and it could be followed by a retention test after four weeks.
- The lack of equipment could be considered as a limitation at different points during the experiment of the main study. The PCs of the school that accepted the invitation to recruit its students for the experiment have integrated graphic cards which affected the VFTG (Island of Volcanoes), and some changes were needed to complete the experiment. Some advanced technical equipment could be used to collect players' data while playing the VFTG, such as an eye tracking system. The Covid-19 lockdown caused the experiment to be cancelled in another secondary school, meaning that the online recruiting to conduct the experiment was considered. However, most of the students did not have suitable laptops to play the VFTG and could not use PCs from friends or public libraries, which led to stopping the online requirement.

In this thesis, new methodologies to design VFTGs have been presented. As such, there are a number of possible extensions to this work which would complement the research of this thesis. The following points have been acknowledged as being of importance in the extension of the presented research in this thesis:

- More prototyping and analysing could be performed to extend and support the current conceptual framework. The proposed framework provides two different skill models (decision-making and teamworking). However, only one skill model was designed and implemented as a part of the Island of Volcanoes VFTG (see Chapter 6, Section 6.3.2.3) due to limitations of time and resources. The teamwork skill model could be prototyped as a next level of the Island of Volcanoes VFTG. In addition, the collected in-game data based on the indicators defined by the evaluation model could be analysed and compared to the data collected after completing the VFTG via pre- and post-tests.
- More studies could be conducted to allow experts to evaluate and provide their perceptions of the framework's connection, usefulness, and usability by inviting a selective small group of experts to a workshop. The framework would be explained in a live session and then the participants would be encouraged to interact with the framework by designing a simple VFTG in a guided session instead of depending on the PowerPoint slides and emails as a way of answering the participants' questions.

- More pedagogical aspects that support and enhance FBL could be modelled in addition to the aspects that the construction of the conceptual framework was based on (ELT, TBL, formative assessment, feedback provision, and social interaction). The construction of the proposed conceptual framework considered feedback provision in ELT modelling, the matching scheme tool, and the evaluation model. Also, social interaction is one of the interaction types utilised in creating the matching scheme tool and skill modelling. This type of interaction can encourage peer feedback. Therefore, a further addition to feedback provision could be achieved by modelling peer feedback and connecting it to game elements and mechanics. Also, the framework could be improved by explicitly handling and predicting the issue of hit-and-trail by players to avoid players would not pay attention to the learning content and simply keep clicking to move to the next task. Currently, the framework provides flexibility to game designers/educators to select a suitable game mechanic for the expected learning outcome based on the task model without a clear guiding to avoid this behaviour.
- Studying immersive technologies such as head-mounted displays (HMDs) and the possible benefits for VFTGs could be an additional extension to this research. HMDs would increase the interaction with virtual environments and the level of presence, which could lead to improved performance. Linking the characteristics of HMDs to FBL would advance the learning experience in VFTGs, especially by enhancing immersion, interaction, visual scanning/observation, and presence.

8.4 Summary

This chapter has summarized the presented research of this thesis, discussed the contributions and how they reflect the thesis aim, and highlighted the limitations and possible future work. This thesis presents a conceptual framework to design VFTGs, which game designers/educators can employ to ensure the effectiveness of VFTGs to serve as an alternative solution when physical Ts are impossible to deliver for any reason. The presented conceptual framework creates connections between learning aspects and game design aspects to facilitate both game mechanics and elements to enhance experiential learning. The conceptual framework includes models (ELT modelling, game modelling, world modelling) in clear steps and each model enriches the design process further.

Appendix A Conceptual Framework for Designing Virtual Field Trip Games Questionnaire

Welcome: this section included a welcome message along with researcher information and privacy and data protection.

Demographic Information

1. To which gender do you most identify:
   - Female
   - Male
   - Intersex
   - Prefer not to say
   - Other

2. Where are you located?

- United Kingdom
- Other

Professional Experience

3. Do you consider yourself as:
   - Educator
   - Game designer
   - Both

4. Years of experience:
   - less than 5 years
   - 5 - 10 years
   - more than 10 years

5. Type of learning institution:
   - School
   - University

6. If you consider your learning institution, you would say it is:
   - A top-level
   - A medium-level
   - A less impressive

7. Considering your duties in a learning institution, would you say that (please mark all appropriate answers):
   - I am involved in developing learning applications.
   - I have administrative/management duties.
   - I have a teaching role.
   - Other

Introduction: this section introduced three important aspects to facilitate understanding the conceptual framework for designing virtual field trip games as following:

- Internal economy: game mechanic.
- Kolb's Experiential Learning theory.

- Game machinations.

The participants provided the option in case they already familiar with any of these aspects to skip and move forward to the next set of questions (Expertise Information).

Expertise Information

8. For each of the following statements, select one box that best describes your agreement regarding your experience with digital games:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I played some digital games.					
I have experience with designing digital games.					
I have used educational digital games in teaching.					
I have experience with designing educational games.					

9. For each of the following statements, select one box that best describes your agreement regarding your experience with field-based learning:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I am familiar with field-based learning.					
I have direct experience with applying field-based learning.					
I am familiar with Experiential Learning Theory.					

10. For each of the following statements, select one box that best describes your agreement regarding your experience with virtual field trip games:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I am familiar with virtual field trip games					
I have direct experience with designing virtual field-trip games.					

The conceptual Framework: this section explained the first version of the framework in general and defines its components which are:
First Link: Experiential Learning Theory (ELT) modelling which maps the theory to the internal economy mechanic (the same as ELT modelling in Section 4.2.1).
Second Link: Matching scheme connects the field-based structure to game elements (the same as Section 4.2.2.1).
Third Link: Environment settings contain four characteristics to transform the field environment into the game environment: Realism, multi-role, multimodal interaction, and complexity. It is more basic and simpler than the world modelling in Section 4.2.3.

The Connection Evaluation

11. For each of the following statements, select one box that best describes your agreement regarding the connections of the three links:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
Kolb's Experiential Learning Theory enhances the design of virtual field trip games.					
The first link connects learning concepts to a game mechanic (internal economy).					
The first link connects learning concepts to game mechanics (internal economy).					
Field-based learning structure of three phases (pre-, during, post-) enhances the design of virtual field trip games.					
Graphical representations (modelling ELT stages as the internal economy) can accurately transpose concepts into practice.					
Graphical representation (the table) can accurately transpose concepts into practice.					
Field-based learning structure of three phases (pre-, during, post-) enhances the design of virtual field trip games.					
The third link connects learning concepts to a game mechanic (internal economy).					
The second link connects learning concepts to game elements.					
The third link can transpose concepts into practice.					

The Usefulness Evaluation

12. For each of the following statements, select one box that best describes your agreement regarding the usefulness of the three links:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The first link is understandable.					
Graphical representations (modelling ELT stages as the internal economy) are explicit.					
The second link is understandable.					
Graphical representation (the table) is explicit.					
The third link is understandable.					

Usability of The Framework:

13. For each of the following statements, mark one box that best describes your reactions to the conceptual framework:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I think that I would like to use this conceptual framework frequently.					
I found the conceptual framework unnecessarily complex.					
I think the conceptual framework would be easy to use.					
I think that I would need some support (examples, more explanation) to be able to use this conceptual framework.					
I found the various steps in this conceptual framework were well integrated.					
I thought there was too much inconsistency in this conceptual framework.					
I would imagine that most experts would learn to use this conceptual framework very quickly.					
I found the conceptual framework very cumbersome to use.					
I would feel very confident when using the conceptual framework.					
I would need to learn a lot of things before I could get going with this conceptual framework.					

Suggestions:

14. For each of the following statements, select one box that best describes your agreement regarding the overall conceptual framework:

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I consider the conceptual framework to be comprehensive of required field-based learning concepts.					
I believe the conceptual framework will help connecting field-based learning to game designing.					
I consider the graphical representations of the framework to be applicable.					

15. Would you consider using the framework in your future research practice?
    - ○ Definitely not
    - ○ Probably not
    - ○ Neutral
    - ○ Probably yes
    - ○ Definitely yes
16. What are the reasons for not applying the framework in your future research?

17. Can you state any missing concepts from the framework that would support designing virtual field trip games?

Appendix B Volcanoes Quiz

Name:

Please circle the correct answer:

1. **The release of gas, lava, and ash from a volcano is called:**
   a) Evacuation.
   b) Eruption.
   c) Epidermis.
2. **Magma and lava are essentially referring to the same thing. The difference between them is that magma is found.......the earth and lava is found........the Earth.**
   a) Outside, inside.
   b) Inside, outside.
   c) Inside, under.
3. **Geologists have detected many small earthquakes in the area near a volcano, what might happen in the near future?**
   a) A volcano may erupt.
   b) A landslide may occur.
   c) A volcano may end.
4. **A town is 20 miles away from an active volcano. The town plans to build a new school and establishes an evacuation plan. Is an evacuation plan the best way to lower the risk to students?**
   a) Yes, the volcano may erupt.
   b) No, the volcano may not erupt.
   c) Yes, a fire may happen.
5. **What type of volcano has alternating layers of lava and ash?**
   a) Shield volcano.
   b) Composite volcano.
   c) Dome volcano.
6. **Volcanoes that have erupted recently are called:**
   a) Extinct
   b) Active
   c) Dormant
7. **Where do volcanic eruptions tend to take place?**
   a) Conservative plate boundaries.
   b) Destructive plate boundaries.
   c) Transform plate boundaries.
8. **What is the method to predict an eruption that can be used by people who live near a volcano and do not have access either to TV or a phone?**
   a) Asking neighbours.
   b) Reading books about that volcano.
   c) Following monitoring alert website.

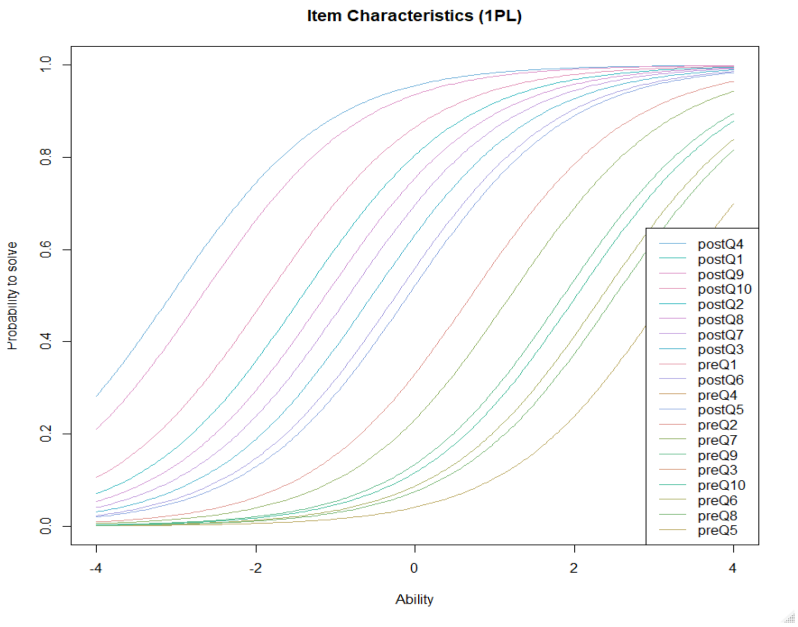
The picture below shows two women running away to escape from the eruption of Mount Sinabung volcano in Indonesia. Use the picture to answer questions 9 and 10.

9. **What volcanic hazard is shown in the picture?**
   a) Lava.
   b) Ash.
   c) Clouds.
10. **What survival item is most needed for those women in the picture?**
   a) Water.
   b) Head wear.
   c) Face mask.

Appendix C Assumption tests of LLTM

Over all data - Rasch model:

```
Call:  RM(X = data[, 4:23])

Conditional log-likelihood: -290.9198
Number of iterations: 10
Number of parameters: 19

Item (Category) Difficulty Parameters (eta): with 0.95 CI:
        Estimate Std. Error lower CI upper CI
preQ2      0.705      0.397   -0.073    1.483
preQ3      1.864      0.411    1.059    2.670
preQ4     -0.240      0.384   -0.993    0.514
preQ5      3.161      0.433    2.312    4.010
preQ6      2.355      0.416    1.541    3.170
preQ7      1.199      0.404    0.408    1.990
preQ8      2.517      0.418    1.698    3.335
preQ9      1.864      0.411    1.059    2.670
preQ10     2.029      0.412    1.221    2.838
postQ1    -3.068      0.468   -3.985   -2.151
postQ2    -1.415      0.383   -2.165   -0.665
postQ3    -0.540      0.381   -1.288    0.207
postQ4    -3.068      0.468   -3.985   -2.151
postQ5    -0.087      0.386   -0.843    0.670
postQ6    -0.240      0.384   -0.993    0.514
postQ7    -0.835      0.380   -1.580   -0.090
postQ8    -1.125      0.380   -1.871   -0.380
postQ9    -2.678      0.435   -3.532   -1.825
postQ10   -1.858      0.393   -2.628   -1.088

Item Easiness Parameters (beta) with 0.95 CI:
             Estimate Std. Error lower CI upper CI
beta preQ1      0.540      0.381   -0.207    1.288
beta preQ2     -0.705      0.397   -1.483    0.073
beta preQ3     -1.864      0.411   -2.670   -1.059
beta preQ4      0.240      0.384   -0.514    0.993
beta preQ5     -3.161      0.433   -4.010   -2.312
beta preQ6     -2.355      0.416   -3.170   -1.541
beta preQ7     -1.199      0.404   -1.990   -0.408
beta preQ8     -2.517      0.418   -3.335   -1.698
beta preQ9     -1.864      0.411   -2.670   -1.059
beta preQ10    -2.029      0.412   -2.838   -1.221
beta postQ1     3.068      0.468    2.151    3.985
beta postQ2     1.415      0.383    0.665    2.165
beta postQ3     0.540      0.381   -0.207    1.288
beta postQ4     3.068      0.468    2.151    3.985
beta postQ5     0.087      0.386   -0.670    0.843
beta postQ6     0.240      0.384   -0.514    0.993
beta postQ7     0.835      0.380    0.090    1.580
beta postQ8     1.125      0.380    0.380    1.871
beta postQ9     2.678      0.435    1.825    3.532
beta postQ10    1.858      0.393    1.088    2.628
```



ICC for items of pre and post measurements

T10 test of equal item difficulties in both subgroups results:

```
Nonparametric RM model test: T10 (global test - subgroup-invariance)
Number of sampled matrices: 1000
Split: median
Group 1: n =  30   Group 2: n = 30
one-sided p-value: 0.29
```

Tpbis test of equal item discrimination for item 1 vs. the other items results:

```
Nonparametric RM model test: Tpbis (discrimination)
    (pointbiserial correlation of test item vs. subscale)
Number of sampled matrices: 1000
Test Item: 1
Subscale  - Items: 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
one-sided p-value (rpbis too low): 0.067
```

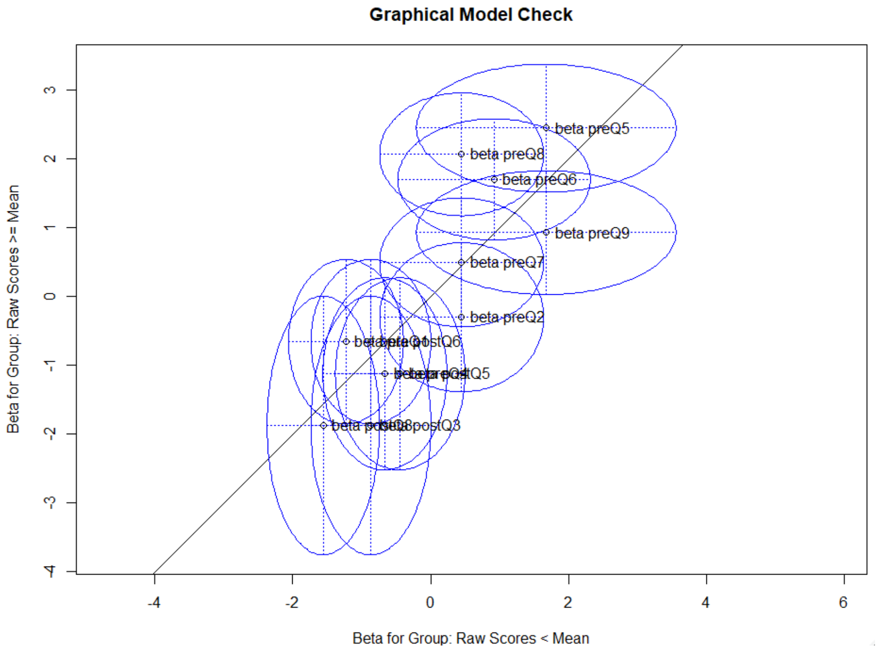
Andersen's (1973) likelihood ratio test with the mean of raw scores as the partitioning criterion:

```
Warning in LRtest.Rm(res.rasch, se = TRUE, splitcr = "median") :

The following items were excluded due to inappropriate response patt
erns within subgroups:
preQ3 preQ10 postQ1 postQ2 postQ4 postQ7 postQ9 postQ10

Full and subgroup models are estimated without these items!
```

Without above-noted items Rasch model fits to data well according to Anderson's LR test if we use split by means or median:

```
Andersen LR-test:
LR-value: 8.53
Chi-square df: 11
p-value:  0.665
```

for 12 remaining items Rasch model is correct (fit well).

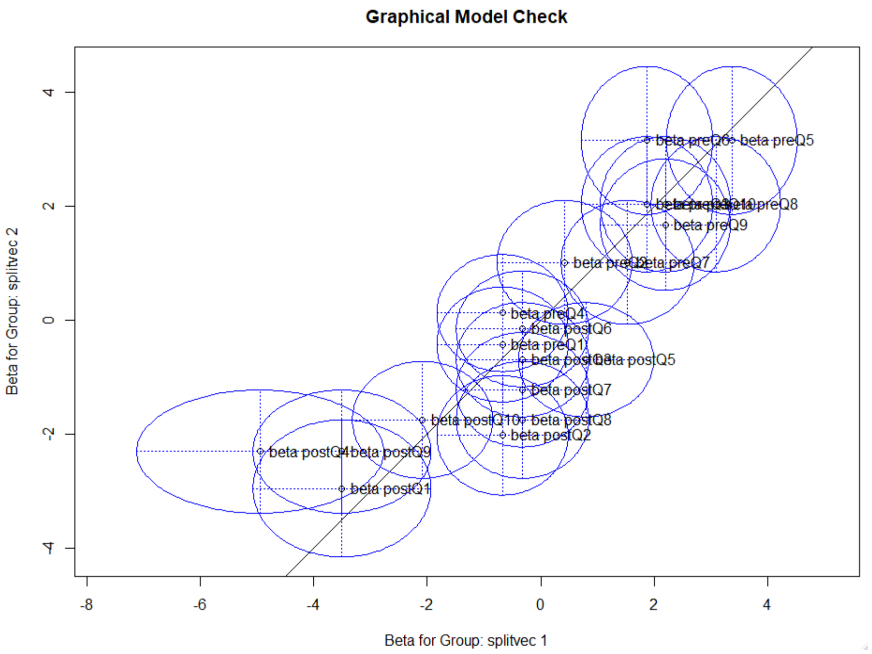


A graphical check of Rasch model fit (splitting by Mean)

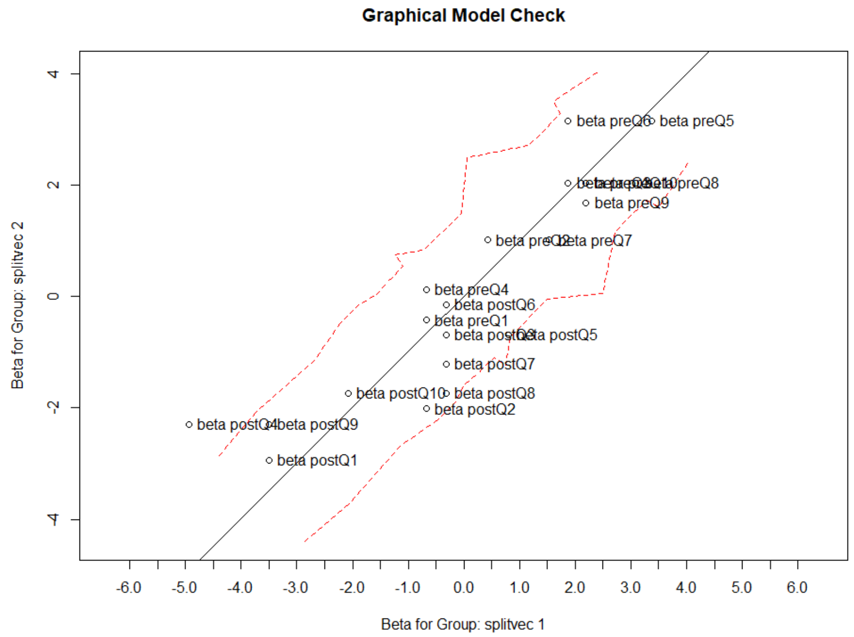
Random splitting for Andersen's (1973) LR test was used (all respondents were randomly assigned to two groups:

```
Andersen LR-test:
LR-value: 23.84
Chi-square df: 19
p-value:  0.202
```

So, Andersen's (1973) LR test indicates that Rasch model fits to data well using random splitting into two groups. As shown in Figure 3, only item 4 confidence interval doesn't overlap with line.



A graphical check of Rasch model fit (Random splitting into 2 groups)



A graphical check of Rasch model fit in another way (Random splitting into 2 groups) – red dotted line is 95% interval boundary

Appendix D Motivation Survey

Name:

Character Name: **Gender**: ☐ Male ☐ Female

1. How often do you play entertaining video games each week?

☐ every day ☐ 3-5 times per week ☐1-2 times per week ☐ Not very often

☐ Not at all

2. (motivation) For each of the following statements, please indicate how true it is for you, using the following scale:

1	2	3	4	5	6	7
not at all true			somewhat true			very true

statement	1	2	3	4	5	6	7
I enjoyed playing this game very much.							
I did not feel nervous at all while playing this game.							
I believe this game could be of some value to me.							
This game was fun to play.							
I thought this was a boring game.							
I felt very tense while playing this game.							
I think that playing this game is useful for learning about volcanos.							
This game did not hold my attention at all.							
I think this is important to play because it can encourage knowing more about volcanoes.							
I would describe this game as very interesting.							
I was very relaxed playing this game.							
I would be willing to do this again because it has some value to me.							
I was anxious while playing this game.							
I thought this game was quite enjoyable.							
I think this is an important game.							
While I was playing this game, I was thinking about how much I enjoyed it.							
I think playing this game could help me to improve my knowledge about volcanos.							
I felt pressured while playing this game.							
I believe playing this game could be beneficial to me.							

Appendix E Presence Survey

2. (presence) Please indicate, whether or not each following statement applies to your experience playing the game. There are no right or wrong answers, only your opinion counts.

In the game world, I had a sense of "being there"								
Not at all	□	□	□	□	□	□	□	**Very much**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
Somehow, I felt that the game world surrounded me.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I felt like I was just perceiving pictures.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I did not feel present in the game space.								
Did not feel	□	□	□	□	□	□	□	**Felt present**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I had a sense of acting in the game space, rather than operating something from outside.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I felt present in the game space.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
How aware were you of the real world surrounding while navigating in the game world? (i.e. sounds, room temperature, other people, etc.)								
Extremely aware	□	□	□	□	□	□	□	**Not aware at all**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I was not aware of my real environment.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I still paid attention to the real environment.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
I was completely captivated by the game world.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
How real did the game world seem to you?								
Completely real	□	□	□	□	□	□	□	**Not real at all**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
How much did your experience in the game environment seem consistent with your real-world experience?								
Not consistent	□	□	□	□	□	□	□	**Very consistent**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	
How real did the game world seem to you?								
About as real as an imagined world	□	□	□	□	□	□	□	Indistinguishable from the real world
	-3	-2	-1	0	1	2	3	
The game world seemed more realistic than the real world.								
Fully disagree	□	□	□	□	□	□	□	**Fully agree**
	-3	**-2**	**-1**	**0**	**1**	**2**	**3**	

References

Abdellatif, A. J., McCollum, B., & McMullan, P. (2018). *Serious games: Quality characteristics evaluation framework and case study.* Paper presented at the Integrated STEM Education Conference (ISEC), 2018 IEEE.

Acquah, E. O., & Katz, H. T. (2020). Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers & Education, 143*, 103667.

Adams, E. (2014). *Fundamentals of game design*: Pearson Education.

Adams, E., & Dormans, J. (2012). *Game mechanics: advanced game design*: New Riders.

Aleman, R., Duball, C., Schwyter, A., & Vaughan, K. (2021). Remote delivery of field experiences in soil sciences. *Natural Sciences Education, 50*(1), e20049.

Alessandri, G., Zuffianò, A., & Perinelli, E. (2017). Evaluating intervention programs with a pretest-posttest design: A structural equation modeling approach. *Frontiers in Psychology, 8*, 223.

Alexiou, A., & Schippers, M. C. (2018). Digital game elements, user experience and learning: A conceptual framework. *Education and Information Technologies*, 1-23.

Amengual Alcover, E., Jaume-i-Capó, A., & Moyà-Alcover, B. (2018). PROGame: A process framework for serious game development for motor rehabilitation therapy. *PloS one, 13*(5), e0197383.

Amory, A. (2007). Game object model version II: a theoretical framework for educational game development. *Educational Technology Research and Development, 55*(1), 51-77.

Anbar, A. D., Mead, C., Bratton III, D., Horodyskyj, L., Hayes, J., Schonstein, D., . . . Leon, A. (2017). Demonstrating the Value of Education Through Exploration as a Theory of Digital Design. *AGUFM, 2017*, ED41B-0275.

Andreoli, R., Corolla, A., Faggiano, A., Malandrino, D., Pirozzi, D., Ranaldi, M., . . . Scarano, V. (2017). A framework to design, develop, and evaluate immersive and collaborative serious games in cultural heritage. *Journal on Computing and Cultural Heritage (JOCCH), 11*(1), 1-22.

Annett, J. (1969). Feedback and human behaviour: The effects of knowledge of results, incentives and reinforcement on learning and performance.

Annetta, L. A. (2010). The “I's” have it: A framework for serious educational game design. *Review of General Psychology, 14*(2), 105.

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M. T. (2009). Investigating the impact of video games on high school students’ engagement and learning about genetics. *Computers & Education, 53*(1), 74-85.

Argles, T., Minocha, S., & Burden, D. (2015). Virtual field teaching has evolved: benefits of a 3D gaming environment. *Geology Today, 31*(6), 222-226.

Arnab, S., Lim, T., Carvalho, M. B., Bellotti, F., Freitas, S., Louchart, S., . . . De Gloria, A. (2015). Mapping learning and game mechanics for serious games analysis. *British journal of educational technology, 46*(2), 391-411.

Arneil, S., Holmes, M., & Street, H. (2001). Review of Hot Potatoes. *Language learning & technology, 5*(2), 28-33.

Ashfield, S., Jarvis, C., & Kaduk, J. (2010). *Serious games for geographical field skills: an exploration.* Paper presented at the Games and Virtual Worlds for Serious Applications (VS-GAMES), 2010 Second International Conference on.

Atchison, C. L., Burmeister, K. C., Egger, A. E., Ryker, K. D., & Tikoff, B. (2020). Designing Remote Field Experiences. Retrieved from

https://nagt.org/nagt/teaching_resources/field/designing_remote_field_experie.html
Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment, 4*(3).
Atwood-Blaine, D., & Huffman, D. (2017). Mobile gaming and student interactions in a science center: the future of gaming in science education. *International journal of science and mathematics education, 15*(1), 45-65.
Australian Government Office for Learning and Teaching. Virtual Field Trips. Retrieved from http://www.virtualfieldtrips.edu.au/
Bandura, A., & Walters, R. H. (1977). *Social learning theory* (Vol. 1): Englewood cliffs Prentice Hall.
Barab, S., Thomas, M., Dodge, T., Carteaux, R., & Tuzun, H. (2005). Making learning fun: Quest Atlantis, a game without guns. *Educational Technology Research and Development, 53*(1), 86-107.
Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.3417*.
Bartholomai, S., & Fitzgerald, C. (2007). The collaborative model of fieldwork education: Implementation of the model in a regional hospital rehabilitation setting. *Australian Occupational Therapy Journal, 54*(s1).
Barzilai, S., & Blau, I. (2014). Scaffolding game-based learning: Impact on learning achievements, perceived learning, and game experiences. *Computers & Education, 70*, 65-79.
Bawa, P., Watson, S. L., & Watson, W. (2018). Motivation is a game: Massively multiplayer online games as agents of motivation in higher education. *Computers & Education, 123*, 174-194.
Behrendt, M., & Franklin, T. (2014). A review of research on school field trips and their value in education. *International Journal of Environmental and Science Education, 9*(3), 235-245.
Bellotti, F., Berta, R., De Gloria, A., & Primavera, L. (2010). Supporting authors in the development of task-based learning in serious virtual worlds. *British journal of educational technology, 41*(1), 86-107.
Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction, 2013*, 1.
Ben-Zadok, G., Leiba, M., & Nachmias, R. (2011). [Chais] Drills, Games or Tests? Evaluating Students' Motivation in Different Online Learning Activities, Using Log File Analysis. *Interdisciplinary Journal of E-Learning and Learning Objects, 7*(1), 235-248.
Bitgood, S. (1989). School field trips: An overview. *Visitor behavior, 4*(2), 3-6.
Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain. New York: David McKay Company. *Inc.(7th Edition 1972)*.
Bloomfield, P. R., & Livingstone, D. (2009). Immersive learning and assessment with quizHUD. *Computing and Information System Journal, 13*(1), 20-26.
Bonello, M. (2001). Fieldwork within the context of higher education: A literature review. *British Journal of Occupational Therapy, 64*(2), 93-99.
Bouchard, S., Robillard, G., St-Jacques, J., Dumoulin, S., Patry, M. J., & Renaud, P. (2004). *Reliability and validity of a single-item measure of presence in VR.* Paper presented at the The 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications.
Bourke, P., & Green, J. (2016). Keeping it real: creating and acquiring assets for serious games. *The Computer Games Journal, 5*(1-2), 7-22.

Bradbeer, J. (1996). Problem-based learning and fieldwork: a better method of preparation? *Journal of Geography in Higher Education, 20*(1), 11-18.

Brinko, K. T. (1993). The practice of giving feedback to improve teaching: What is effective? *The Journal of Higher Education, 64*(5), 574-593.

Brom, C., Buchtová, M., Šisler, V., Děchtěrenko, F., Palme, R., & Glenk, L. M. (2014). Flow, social interaction anxiety and salivary cortisol responses in serious games: A quasi-experimental study. *Computers & Education, 79*, 69-100.

Brooke, J. (1996). SUS-A quick and dirty usability scale. *Usability evaluation in industry, 189*(194), 4-7.

Brookhart, S. M. (2008). *How to give effective feedback to your students*: ASCD.

Burden, D., Argles, T., Minocha, S., Rock, J., Tilling, S., & Wheeler, P. (2017). Fieldscapes–Creating and Evaluating a 3D Virtual Field Trip System.

Burke, J., McNeill, M., Charles, D., Morrow, P., Crosbie, J., & McDonough, S. (2010). *Designing engaging, playable games for rehabilitation.* Paper presented at the Proceedings of the 8th international conference on disability, virtual reality & associated technologies.

Bursztyn, N., Walker, A., Shelton, B., & Pederson, J. (2017). Assessment of student learning using augmented reality Grand Canyon field trips for mobile smart devices. *Geosphere, 13*(2), 260-268.

Calvert, J., & Abadia, R. (2020). Impact of immersing university and high school students in educational linear narratives using virtual reality technology. *Computers & Education, 159*, 104005.

Calvo-Morata, A., Alonso-Fernández, C., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2020). Serious games to prevent and detect bullying and cyberbullying: A systematic serious games and literature review. *Computers & Education, 157*, 103958.

Camerer, C. (1998). Bounded rationality in individual decision making. *Experimental economics, 1*(2), 163-183.

Carvalho, M. B., Bellotti, F., Berta, R., De Gloria, A., Sedano, C. I., Hauge, J. B., . . . Rauterberg, M. (2015). An activity theory-based model for serious games analysis and conceptual design. *Computers & Education, 87*, 166-181.

Chan, C. S., Chan, Y.-h., & Fong, T. H. A. (2020). Game-based e-learning for urban tourism education through an online scenario game. *International Research in Geographical and Environmental Education, 29*(4), 283-300.

Chang, K. E., Wu, L.-J., Weng, S. E., & Sung, Y. T. (2012). Embedding game-based problem-solving phase into problem-posing system for mathematics learning. *Computers & Education, 58*(2), 775-786.

Charsky, D. (2010). From edutainment to serious games: A change in the use of game characteristics. *Games and culture, 5*(2), 177-198.

Chen, C. h. (2010). The implementation and evaluation of a mobile self-and peer-assessment system. *Computers & Education, 55*(1), 229-236.

Chen, C. H., Liu, G. Z., & Hwang, G. J. (2016). Interaction between gaming and multistage guiding strategies on students' field trip mobile learning performance and motivation. *British journal of educational technology, 47*(6), 1032-1050.

Chen, S., & Michael, D. (2005). Proof of learning: Assessment in serious games. *Retrieved October, 17*, 2008.

Chen, Y. L., & Hsu, C. C. (2020). Self-regulated mobile game-based English learning in a virtual reality environment. *Computers & Education, 154*, 103910.

Chen, Z. H., Chen, H. H. J., & Dai, W. J. (2018). Using narrative-based contextual games to enhance language learning: A case study. *Journal of Educational Technology & Society, 21*(3), 186-198.

Cheng, K. H. (2021). Teachers' perceptions of exploiting immersive virtual field trips for learning in primary education. *Journal of Research on Technology in Education*, 1-18.

Cheng, K. H., & Tsai, C. C. (2019). A case study of immersive virtual field trips in an elementary classroom: Students' learning experience and teacher-student interaction behaviors. *Computers & Education, 140*, 103600.

Chew, B. S. (2017). *An efficient framework for game-based learning activity.* Paper presented at the Teaching, Assessment, and Learning for Engineering (TALE), 2017 IEEE 6th International Conference on.

Chin, K. Y., & Wang, C. S. (2021). Effects of augmented reality technology in a mobile touring system on university students' learning performance and interest. *Australasian Journal of Educational Technology, 37*(1), 27-42.

Clark, D., Nelson, B., Sengupta, P., & D'Angelo, C. (2009). *Rethinking science learning through digital games and simulations: Genres, examples, and evidence.* Paper presented at the Learning science: Computer games, simulations, and education workshop sponsored by the National Academy of Sciences, Washington, DC.

Coenen, T., Mostmans, L., & Naessens, K. (2013). MuseUs: Case study of a pervasive cultural heritage serious game. *Journal on Computing and Cultural Heritage (JOCCH), 6*(2), 1-19.

Cohn, E. S., Dooley, N. R., & Simmons, L. A. (2002). Collaborative learning applied to fieldwork education. *Occupational therapy in health care, 15*(1-2), 69-83.

Collins, T., Gaved, M., & Lea, J. (2010). Remote fieldwork: using portable wireless networks and backhaul links to participate remotely in fieldwork.

Conejo, R., Guzmán, E., Millán, E., Trella, M., Pérez-De-La-Cruz, J. L., & Ríos, A. (2004). SIETTE: A web-based tool for adaptive testing. *International Journal of Artificial Intelligence in Education, 14*(1), 29-61.

Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J, 13*(1), 17-31.

Csikszentmihalyi, M., & Csikszentmihalyi, I. (1975). *Beyond boredom and anxiety* (Vol. 721): Jossey-Bass San Francisco.

Czub, M., & Piskorz, J. (2014). *How body movement influences virtual reality analgesia?* Paper presented at the 2014 International Conference on Interactive Technologies and Games.

Davidson, S. K., Passmore, C., & Anderson, D. (2010). Learning on zoo field trips: The interaction of the agendas and practices of students, teachers, and zoo educators. *Science Education, 94*(1), 122-141.

de-Marcos, L., Hilera, J. R., Barchino, R., Jiménez, L., Martínez, J. J., Gutiérrez, J. A., . . . Otón, S. (2010). An experiment for improving students performance in secondary and tertiary education by means of m-learning auto-assessment. *Computers & Education, 55*(3), 1069-1079.

De Freitas, S., & Jarvis, S. (2006). A framework for developing serious games to meet learner needs.

de Paz-Álvarez, M. I., Blenkinsop, T. G., Buchs, D. M., Gibbons, G. E., & Cherns, L. (2021). Virtual fieldtrip to the Esla Nappe (Cantabrian Zone, NW Spain): delivering traditional geological mapping skills remotely using real data. *Solid Earth*.

Denis, G., & Jouvelot, P. (2005). *Motivation-driven educational game design: applying best practices to music education.* Paper presented at the Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology.

Denton, P., Madden, J., Roberts, M., & Rowe, P. (2008). Students' response to traditional and computer-assisted formative feedback: A comparative case study. *British journal of educational technology, 39*(3), 486-500.

Derbali, L., & Frasson, C. (2012). Assessment of learners' motivation during interactions with serious games: A study of some motivational strategies in food-force. *Advances in Human-Computer Interaction, 2012*.

DeWitt, J., & Storksdieck, M. (2008). A short review of school field trips: Key findings from the past and implications for the future. *Visitor Studies, 11*(2), 181-197.

Dhir, S. K., Verma, D., Batta, M., & Mishra, D. (2017). E-learning in medical education in India. *Indian pediatrics, 54*(10), 871-877.

Dieleman, H., & Huisingh, D. (2006). Games by which to learn and teach about sustainable development: exploring the relevance of games and experiential learning for sustainability. *Journal of Cleaner Production, 14*(9-11), 837-847.

Dimitrov, D. M., & Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work, 20*(2), 159-165.

Discovery Education. Discovery Education Virtual Field Trips. Retrieved from http://www.discoveryeducation.com/Events/virtual-field-trips/explore/index.cfm?campaign=flyout_parents_virtual_field_trips

Divjak, B., & Tomić, D. (2011). The impact of game-based learning on the achievement of learning goals and motivation for learning mathematics-literature review. *Journal of Information and Organizational Sciences, 35*(1), 15-30.

Djaouti, D., Alvarez, J., & Jessel, J. P. (2011). Classifying serious games: the G/P/S model. *Handbook of research on improving learning and motivation through educational games: Multidisciplinary approaches, 2*, 118-136.

Djaouti, D., Alvarez, J., Jessel, J. P., Methel, G., & Molinier, P. (2008). A gameplay definition through videogame classification. *International Journal of Computer Games Technology, 2008*, 4.

Djonko-Moore, C. M., & Joseph, N. M. (2016). Out of the classroom and into the city: The use of field trips as an experiential learning tool in teacher education. *Sage Open, 6*(2), 2158244016649648.

Dolphin, G., Dutchak, A., Karchewski, B., & Cooper, J. (2019). Virtual field experiences in introductory geology: Addressing a capacity problem, but finding a pedagogical one. *Journal of Geoscience Education, 67*(2), 114-130.

dos Santos Nunes, E. P., Roque, L. G., & dos Santos Nunes, F. d. L. (2016). Measuring Knowledge Acquisition in 3D Virtual Learning Environments. *IEEE computer graphics and applications, 36*(2), 58-67.

Dreyfus, S. E., & Dreyfus, H. L. (1980). *A five-stage model of the mental activities involved in directed skill acquisition*. Retrieved from

Dringus, L. P., & Ellis, T. (2005). Using data mining as a strategy for assessing asynchronous discussion forums. *Computers & Education, 45*(1), 141-160.

Dummer, T. J., Cook, I. G., Parker, S. L., Barrett, G. A., & Hull, A. P. (2008). Promoting and assessing 'deep learning'in geography fieldwork: An evaluation of reflective field diaries. *Journal of Geography in Higher Education, 32*(3), 459-479.

Dunphy, A., & Spellman, G. (2009). Geography fieldwork, fieldwork value and learning styles. *International Research in Geographical and Environmental Education, 18*(1), 19-28.

Dyson, L. E., Lawrence, E., Litchfield, A., & Zmijewska, A. (2008). *M-fieldwork for information systems students.* Paper presented at the Hawaii International Conference on System Sciences, Proceedings of the 41st Annual.

Edwards, S. H. (2003). *Using test-driven development in the classroom: Providing students with automatic, concrete feedback on performance.* Paper presented at the Proceedings of the international conference on education and information systems: technologies and applications EISTA.

Efstathiou, I., Kyza, E. A., & Georgiou, Y. (2018). An inquiry-based augmented reality mobile learning approach to fostering primary school students' historical reasoning in non-formal settings. *Interactive Learning Environments, 26*(1), 22-41.

Eseryel, D., Guo, Y., & Law, V. (2012). Interactivity 3 Design and Assessment Framework for Educational Games to Promote Motivation and Complex Problem-Solving Skills. In *Assessment in game-based learning* (pp. 257-285): Springer.

Eseryel, D., Law, V., Ifenthaler, D., Ge, X., & Miller, R. (2014). An investigation of the interrelationships between motivation, engagement, and complex problem solving in game-based learning. *Educational technology & society, 17*(1), 42-53.

Espasa, A., & Meneses, J. (2010). Analysing feedback processes in an online teaching and learning environment: an exploratory study. *Higher Education, 59*(3), 277-292.

Farrell, D., & Moffat, D. (2014). *Applying the self determination theory of motivation in games based learning.* Paper presented at the European Conference on Games Based Learning.

Faulconer, E. K., Griffith, J., Wood, B., Acharyya, S., & Roberts, D. (2018). A comparison of online, video synchronous, and traditional learning modes for an introductory undergraduate physics course. *Journal of Science Education and Technology, 27*(5), 404-411.

Fedesco, H. N., Cavin, D., & Henares, R. (2020). Field-Based Learning in Higher Education: Exploring the Benefits and Possibilities. *Journal of the Scholarship of Teaching and Learning, 20*(1), 65-84.

Felnhofer, A., Hetterle, T., Schmidt, M., Kryspin-Exner, I., & Kothgassner, O. (2014). Challenging gender differences in physical presence: Affective states, spatial abilities and presence in a relaxing virtual park. *Challenging Presence. Proceedings of the International Society for Presence Research*, 161-166.

Felnhofer, A., Kothgassner, O. D., Beutl, L., Hlavacs, H., & Kryspin-Exner, I. (2012). Is virtual reality made for men only? Exploring gender differences in the sense of presence. *Proceedings of the International Society on presence research*, 103-112.

Feng, Z., González, V. A., Amor, R., Lovreglio, R., & Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: A systematic literature review. *Computers & Education, 127*, 252-266.

Finn, J. D., & Rock, D. A. (1997). Academic success among students at risk for school failure. *Journal of applied psychology, 82*(2), 221.

Fischer, G. H. (1995). Derivations of the Rasch model. In *Rasch models* (pp. 15-38): Springer.

Fischer, G. H., & Molenaar, I. W. (1995). *Rasch Models: Foundations, Recent Developments, and Applications*: Springer Science & Business Media.

Fjællingsdal, K. S., & Klöckner, C. A. (2017). ENED-GEM: A Conceptual Framework Model for Psychological Enjoyment Factors and Learning Mechanisms in Educational Games about the Environment. *Frontiers in Psychology, 8*, 1085.

Fowler, A. (2013). *Measuring learning and fun in video games for young children: a proposed method.* Paper presented at the Proceedings of the 12th international conference on interaction design and children.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research, 74*(1), 59-109.

Fredricks, J. A., & Eccles, J. S. (2006). Is extracurricular participation associated with beneficial outcomes? Concurrent and longitudinal relations. *Developmental psychology, 42*(4), 698.

Frontczak, N. T. (1998). A paradigm for the selection, use and development of experiential learning activities in marketing education. *Marketing Education Review, 8*(3), 25-33.

Fu, Q. K., Lin, C. J., Hwang, G. J., & Zhang, L. (2019). Impacts of a mind mapping-based contextual gaming approach on EFL students' writing performance, learning perceptions and generative uses in an English course. *Computers & Education, 137*, 59-77.

Gaber, J. (2007). Simulating planning: SimCity as a pedagogical tool. *Journal of Planning Education and Research, 27*(2), 113-121.

Garneli, V., Giannakos, M., & Chorianopoulos, K. (2017). Serious games as a malleable learning medium: The effects of narrative, gameplay, and making on students' performance and attitudes. *British journal of educational technology, 48*(3), 842-859.

Georgiadi, N., Kokkoli-Papadopoulou, E., Kordatos, G., Partheniadis, K., Sparakis, M., Koutsabasis, P., . . . Stavrakis, M. (2016). *A pervasive role-playing game for introducing elementary school students to archaeology.* Paper presented at the Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct.

Gerhard, M., Moore, D. J., & Hobbs, D. J. (2001). *Continuous presence in collaborative virtual environments: Towards a hybrid avatar-agent model for user representation.* Paper presented at the International Workshop on Intelligent Virtual Agents.

Getchell, K., Miller, A., Nicoll, J. R., Sweetman, R., & Allison, C. (2010). Games methodologies and immersive environments for virtual fieldwork. *IEEE Transactions on Learning Technologies, 3*(4), 281-293.

Ghergulescu, I., & Muntean, C. H. (2012). Measurement and analysis of learner's motivation in game-based e-learning. In *Assessment in game-based learning* (pp. 355-378): Springer.

Graafland, M., Dankbaar, M., Mert, A., Lagro, J., De Wit-Zuurendonk, L., Schuit, S., . . . Schijven, M. (2014). How to systematically assess serious games applied to health care. *JMIR Serious Games, 2*(2).

Grassini, S., & Laumann, K. (2020). Questionnaire measures and physiological correlates of presence: A systematic review. *Frontiers in Psychology, 11*, 349.

Green, D. R., & Bojar, K. (2010). "Ythanview"—Visualizing an Estuary and Virtual Fieldwork at the Ythan Estuary, Scotland, UK. In *Coastal and Marine Geospatial Technologies* (pp. 171-186): Springer.

Greene, J. P., Kisida, B., & Bowen, D. H. (2014). The educational value of field trips. *Education Next, 14*(1).

Gregory, D. D., Tomes, H. E., Panasiuk, S. L., & Andersen, A. J. (2022). Building an online field course using digital and physical tools including VR field sites and virtual core logging. *Journal of Geoscience Education, 70*(1), 85-100.

Gregory, R., & Long, G. (2009). Using structured decision making to help implement a precautionary approach to endangered species management. *Risk Analysis: An International Journal, 29*(4), 518-532.

Gutwill, J. P., & Allen, S. (2012). Deepening students' scientific inquiry skills during a science museum field trip. *Journal of the Learning Sciences, 21*(1), 130-181.

Hajirasouli, A., Banihashemi, S., Kumarasuriyar, A., Talebi, S., & Tabadkani, A. (2021). Virtual reality-based digitisation for endangered heritage sites: Theoretical framework and application. *Journal of Cultural Heritage, 49*, 140-151.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior, 54*, 170-179.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*: Sage.

Han, I. (2020). Immersive virtual field trips in education: A mixed-methods study on elementary students' presence and perceived learning. *British journal of educational technology, 51*(2), 420-435.

Harrington, M. C. (2009). An ethnographic comparison of real and virtual reality field trips to Trillium Trail: the salamander find as a salient event. *Children Youth and Environments, 19*(1), 74-101.

Harrington, M. C., Bledsoe, Z., Jones, C., Miller, J., & Pring, T. (2021). Designing a Virtual Arboretum as an Immersive, Multimodal, Interactive, Data Visualization Virtual Field Trip. *Multimodal Technologies and Interaction, 5*(4), 18.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research, 77*(1), 81-112.

Hauge, J. B., Barenbrock, A., & Thoben, K.-D. (2017). *Avatars as motivational factor in simulation games.* Paper presented at the The International Scientific Conference eLearning and Software for Education.

Hautala, J., Heikkilä, R., Nieminen, L., Rantanen, V., Latvala, J. M., & Richardson, U. (2020). Identification of reading difficulties by a digital game-based assessment technology. *Journal of Educational Computing Research*, 0735633120905309.

Healey, M., & Jenkins, A. (2000). Kolb's experiential learning theory and its application in geography in higher education. *Journal of geography, 99*(5), 185-195.

Hein, D., Mai, C., & Hußmann, H. (2018). *The usage of presence measurements in research: a review.* Paper presented at the Proceedings of the International Society for Presence Research Annual Conference (Presence'18). The International Society for Presence Research.

Heracleous, L. T. (1994). Rational decision making: myth or reality? *Management development review*.

Hershkovitz, A., & Nachmias, R. (2008). *Developing a Log-based Motivation Measuring Tool.* Paper presented at the EDM.

Higgins, N., Dewhurst, E., & Watkins, L. (2012). Field trips as short-term experiential learning activities in legal education. *The Law Teacher, 46*(2), 165-178.

Hill, J., & Woodland, W. (2002). An evaluation of foreign fieldwork in promoting deep learning: A preliminary investigation. *Assessment & Evaluation in Higher Education, 27*(6), 539-555.

Hobbs, L., Stevens, C., Hartley, J., & Hartley, C. (2019). Science Hunters: an inclusive approach to engaging with science through Minecraft. *JCOM: Journal of Science Communication, 18*(2).

Hoffman, B., & Nadelson, L. (2010). Motivational engagement and video gaming: A mixed methods study. *Educational Technology Research and Development, 58*(3), 245-270.

Horodyskyj, L. B., Mead, C., Belinson, Z., Buxner, S., Semken, S., & Anbar, A. D. (2018). Habitable Worlds: Delivering on the promises of online education. *Astrobiology, 18*(1), 86-99.

Hsu, T. Y., & Chen, C. M. (2010). A mobile learning module for high school fieldwork. *Journal of Geography, 109*(4), 141-149.

Hunicke, R., LeBlanc, M., & Zubek, R. (2004). *MDA: A formal approach to game design and game research.* Paper presented at the Proceedings of the AAAI Workshop on Challenges in Game AI.

Hwang, G. J., & Chang, H. F. (2011). A formative assessment-based mobile learning approach to improving the learning attitudes and achievements of students. *Computers & Education, 56*(4), 1023-1031.

Hwang, G. J., Hsiao, C. L., & Tseng, J. C. (2003). A computer-assisted approach to diagnosing student learning problems in science courses. *J. Inf. Sci. Eng., 19*(2), 229-248.

Ifenthaler, D., Eseryel, D., & Ge, X. (2012). Assessment for game-based learning. In *Assessment in game-based learning* (pp. 1-8): Springer.

Ihantola, P., Ahoniemi, T., Karavirta, V., & Seppälä, O. (2010). *Review of recent systems for automatic assessment of programming assignments.* Paper presented at the

Proceedings of the 10th Koli Calling International Conference on Computing Education Research.

IPQ. (1999). Igroup presence questionnaire. Retrieved from http://www.igroup.org/pq/ipq/index.php

Jacobson, A. R., Militello, R., & Baveye, P. C. (2009). Development of computer-assisted virtual field trips to support multidisciplinary learning. *Computers & Education, 52*(3), 571-580.

Jordan, S., & Mitchell, T. (2009). e-Assessment for learning? The potential of short-answer free-text questions with tailored feedback. *British journal of educational technology, 40*(2), 371-385.

Jose, S., Patrick, P. G., & Moseley, C. (2017). Experiential learning theory: the importance of outdoor classrooms in environmental education. *International Journal of Science Education, Part B, 7*(3), 269-284.

Jung, B., Sainsbury, S., Grum, R. M., Wilkins, S., & Tryssenaar, J. (2002). Collaborative fieldwork education with student occupational therapists and student occupational therapist assistants. *Canadian Journal of Occupational Therapy, 69*(2), 95-103.

Kamarainen, A., Reilly, J., Metcalf, S., Grotzer, T., & Dede, C. (2018). Using mobile location-based augmented reality to support outdoor learning in undergraduate ecology and environmental science courses. *Bulletin of the Ecological Society of America, 99*(2), 259-276.

Kamarainen, A. M., Metcalf, S., Grotzer, T., Browne, A., Mazzuca, D., Tutwiler, M. S., & Dede, C. (2013). EcoMOBILE: Integrating augmented reality and probeware with environmental education field trips. *Computers & Education, 68*, 545-556.

Karadeniz, Ş. (2009). The impacts of paper, web and mobile based assessment on students achievement and perceptions. *Scientific Research and Essays, 4*(10), 984-991.

Kayalar, C., & Balcısoy, S. (2008). Augmented reality based user interfaces to assist fieldwork on excavation sites.

Kelly, H., Howell, K., Glinert, E., Holding, L., Swain, C., Burrowbridge, A., & Roper, M. (2007). How to build serious games. *Communications of the ACM, 50*(7), 44-49.

Kenna, J. L., & Potter, S. (2018). Experiencing the world from inside the classroom: Using virtual field trips to enhance social studies instruction. *The Social Studies, 109*(5), 265-275.

Kent, M., Gilbertson, D. D., & Hunt, C. O. (1997). Fieldwork in geography teaching: A critical review of the literature and approaches. *Journal of Geography in Higher Education, 21*(3), 313-332.

Kiili, K. (2005). Digital game-based learning: Towards an experiential gaming model. *The Internet and higher education, 8*(1), 13-24.

Kiili, K. (2007). Foundation for problem-based gaming. *British journal of educational technology, 38*(3), 394-404.

Kim, H., Ke, F., & Paek, I. (2017). Game-based learning in an OpenSim-supported virtual environment on perceived motivational quality of learning. *Technology, Pedagogy and Education, 26*(5), 617-631.

Kim, S., Chang, M., Deater-Deckard, K., Evans, M. A., Norton, A., & Samur, Y. (2017). Educational games and students' game engagement in elementary school classrooms. *Journal of Computers in Education, 4*(4), 395-418.

Kingston, D., Eastwood, W., Jones, P., Johnson, R., Marshall, S., & Hannah, D. (2012). Experiences of using mobile technologies and virtual field tours in Physical Geography: implications for hydrology education. *Hydrology and Earth System Sciences, 16*(5), 1281-1286.

Klippel, A., Zhao, J., Jackson, K. L., La Femina, P., Stubbs, C., Wetzel, R., . . . Oprean, D. (2019). Transforming earth science education through immersive experiences: Delivering on a long held promise. *Journal of Educational Computing Research, 57*(7), 1745-1771.

Klippel, A., Zhao, J., Oprean, D., Wallgrün, J. O., & Chang, J. S. K. (2019). *Research framework for immersive virtual field trips.* Paper presented at the 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).

Klippel, A., Zhao, J., Oprean, D., Wallgrün, J. O., Stubbs, C., La Femina, P., & Jackson, K. L. (2020). The value of being there: Toward a science of immersive virtual field trips. *Virtual Reality, 24*(4), 753-770.

Klippel, A., Zhao, J., Sajjadi, P., Wallgrün, J. O., Bagher, M. M., & Oprean, D. (2020). *Immersive Place-based Learning–An Extended Research Framework.* Paper presented at the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).

Kluger, A. N., & DeNisi, A.The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. (1996). American Psychological Association.

Kolb, D. A.Experimental Learning: Experience as the source of learning and development. (1984). Prentice Hall: Englewood Cliffs, NJ.

Kolb, D. A. (1993). The process of experiential learning. *Culture and processes of adult learning*, 138-156.

Kremer, K. E. (2012). Conducting game user experience research with preschoolers. *May 2012, Austin, Texas, USA, 307*.

Lachlan, K., & Krcmar, M. (2011). Experiencing presence in video games: The role of presence tendencies, game experience, gender, and time spent in play. *Communication Research Reports, 28*(1), 27-31.

Ladyshewsky, R. K., & Gardner, P. (2008). Peer assisted learning and blogging: A strategy to promote reflective practice during clinical fieldwork. *Australasian Journal of Educational Technology, 24*(3).

LEGO DC Supervillains. (Standard Edition)[Video game]. (2018). USA: Interactive Entertainment.

LEGO Marvel Super Heroes 2. (Standard Edition)[Video game]. (2017). USA: Interactive Entertainment.

Lenkeit Meezan, K. A., & Cuffey, K. (2012). Virtual field trips for introductory geoscience classes.

Lessiter, J., Freeman, J., Keogh, E., & Davidoff, J. (2001). A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence: Teleoperators & Virtual Environments, 10*(3), 282-297.

Li, M. C., & Tsai, C. C. (2013). Game-based learning in science education: A review of relevant research. *Journal of Science Education and Technology, 22*(6), 877-898.

Liao, C. W., Chen, C. H., & Shih, S. J. (2019). The interactivity of video and collaboration for learning achievement, intrinsic motivation, cognitive load, and behavior patterns in a digital game-based learning environment. *Computers & Education, 133*, 43-55.

Lieberoth, A., Pedersen, M. K., & Sherson, J. (2015). Play or science?: a study of learning and framing in crowdscience games. *arXiv preprint arXiv:1510.06841*.

Lilley, M., Barker, T., & Britton, C. (2004). The development and evaluation of a software prototype for computer-adaptive testing. *Computers & Education, 43*(1), 109-123.

Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. *Chae, S.-Kang, U.–Jeon, E.–Linacre, JM (eds.): Development of Computerised Middle School Achievement Tests, MESA Research Memorandum*(69).

Liu, C., Agrawal, P., Sarkar, N., & Chen, S. (2009). Dynamic difficulty adjustment in computer games through real-time anxiety-based affective feedback. *International Journal of Human-Computer Interaction, 25*(6), 506-529.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of computer-mediated communication, 3*(2), JCMC321.

Lombard, M., Ditton, T. B., & Weinstein, L. (2009). *Measuring presence: the temple presence inventory.* Paper presented at the Proceedings of the 12th annual international workshop on presence.

Lonergan, N., & Andresen, L. W. (1988). Field-based education: Some theoretical considerations. *Higher education research and development, 7*(1), 63-77.

Luo, L., Yin, H., Cai, W., Zhong, J., & Lees, M. (2017). Design and evaluation of a data-driven scenario generation framework for game-based training. *IEEE Transactions on Computational Intelligence and AI in Games, 9*(3), 213-226.

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R.

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive science, 5*(4), 333-369.

Mani, L., Cole, P. D., & Stewart, I. (2016). Using video games for volcanic hazard education and communication: an assessment of the method and preliminary results.

Marougkas, A., Troussas, C., Krouska, A., & Sgouropoulou, C. (2021). A Framework for Personalized Fully Immersive Virtual Reality Learning Environments with Gamified Design in Education. In *Novelties in Intelligent Digital Systems* (pp. 95-104): IOS Press.

Marshall, M. S., & Higley, M. C. (2021). Multi-scale virtual field experience: sedimentology and stratigraphy of Grand Ledge, Michigan, USA. *Geoscience Communication, 4*(4), 461-474.

Martín-Hernández, P., Gil-Lacruz, M., Gil-Lacruz, A. I., Azkue-Beteta, J. L., Lira, E. M., & Cantarero, L. (2021). Fostering University Students' Engagement in Teamwork and Innovation Behaviors through Game-Based Learning (GBL). *Sustainability, 13*(24), 13573.

Mason, L. (1998). Fieldwork education: Collaborative group learning in community settings. *Australian Occupational Therapy Journal, 45*(4), 124-130.

Mathews, S., Andrews, L., & Luck, E. (2012). Developing a Second Life virtual field trip for university students: an action research approach. *Educational Research, 54*(1), 17-38.

McAuley, E., Duncan, T., & Tammen, V. V. (1989). Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport, 60*(1), 48-58.

McGowan, E. G., & Scarlett, J. P. (2021). Volcanoes in video games: the portrayal of volcanoes in commercial off-the-shelf (COTS) video games and their learning potential. *Geoscience Communication, 4*(1), 11-31.

McGuinness, M., & Simm, D. (2005). Going global? Long-haul fieldwork in undergraduate geography. *Journal of Geography in Higher Education, 29*(2), 241-253.

Mead, C., Buxner, S., Bruce, G., Taylor, W., Semken, S., & Anbar, A. D. (2019). Immersive, interactive virtual field trips promote science learning. *Journal of Geoscience Education, 67*(2), 131-142.

Mekler, E. D., Brühlmann, F., Opwis, K., & Tuch, A. N. (2013). *Do points, levels and leaderboards harm intrinsic motivation?: an empirical analysis of common gamification elements.* Paper presented at the Proceedings of the First International Conference on gameful design, research, and applications.

Melo, M., Raposo, J., Coelho, A., Narciso, D., & Bessa, M. (2019). Immersive 360° video user experience: impact of different variables in the sense of presence and cybersickness.

Melo, M., Sampaio, S., Barbosa, L., Raposo, J., & Bessa, M. (2016). *The impact of different exposure times to 360 video experience on the sense of presence.* Paper presented at the 2016 23rd Portuguese Meeting on Computer Graphics and Interaction (EPCGI).

Miljanovic, M. A., & Bradbury, J. S. (2018). *Making Serious Programming Games Adaptive.* Paper presented at the Joint International Conference on Serious Games.

Miller, L., Moreno, J., Willcockson, I., Smith, D., & Mayes, J. (2006). An online, interactive approach to teaching neuroscience to adolescents. *CBE-Life Sciences Education, 5*(2), 137-143.

Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education, 34*(2), 181-192.

Min, W., Frankosky, M., Mott, B. W., Rowe, J., Smith, P. A. M., Wiebe, E., . . . Lester, J. (2019). DeepStealth: game-based learning stealth assessment with deep neural networks. *IEEE Transactions on Learning Technologies*.

Minaei-Bidgoli, B., & Punch, W. (2003). *Using genetic algorithms for data mining optimization in an educational web-based system.* Paper presented at the Genetic and Evolutionary Computation—GECCO 2003.

Mitgutsch, K., & Alvarado, N. (2012). *Purposeful by design?: a serious game design assessment framework.* Paper presented at the Proceedings of the International Conference on the foundations of digital games.

Mogk, D. W., & Goodwin, C. (2012). Learning in the field: Synthesis of research on thinking and learning in the geosciences. *Geological society of America special papers, 486*(0), 131-163.

Molan, S., Weber, D., & Kor, M. (2022). Shaping Children's Knowledge and Response to Bushfire Through Use of an Immersive Virtual Learning Environment. *Journal of Educational Computing Research*, 07356331211054569.

Monster Hunter Generations: Ultimate. (Standard Edition)[Video game]. (2018). Japan: Capcom, Osaka.

Mory, E. H. (2004). Feedback research revisited. *Handbook of research on educational communications and technology, 2*, 745-783.

Mulders, M., Buchner, J., & Kerres, M. (2020). A framework for the use of immersive virtual reality in learning environments. *International Journal of Emerging Technologies in Learning (iJET), 15*(24), 208-224.

Nadolny, L., Valai, A., Cherrez, N. J., Elrick, D., Lovett, A., & Nowatzke, M. (2020). Examining the characteristics of game-based learning: A content analysis and design framework. *Computers & Education, 156*, 103936.

Narciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction, 16*(4), 310-322.

Needle, M. D., Crider, J. G., Mooc, J., & Akers, J. F. (2021). Virtual field experiences in a web-based videogame environment: Open-ended examples of existing and fictional field sites. *Geoscience Communication Discussions*, 1-13.

Neuwirth, C. (2020). EQUILIBRIUM game: a virtual field trip through a complex system. *System Dynamics Review, 36*(2), 223-238.

Nicol, D. (2007). *Principles of good assessment and feedback: Theory and practice.* Paper presented at the REAP International online conference on assessment design for learner responsibility.

Nicovich, S. G., Boller, G. W., & Cornwell, T. B. (2005). Experienced presence within computer-mediated communications: Initial explorations on the effects of gender with respect to empathy and immersion. *Journal of computer-mediated communication, 10*(2), JCMC1023.

Nieuwhof-Leppink, A., de Jong, T., van de Putte, E., & Schappin, R. (2019). Does a serious game increase intrinsic motivation in children receiving urotherapy? *Journal of pediatric urology, 15*(1), 36. e31-36. e37.

Ninaus, M., Kiili, K., McMullen, J., & Moeller, K. (2017). Assessing fraction knowledge by a digital game. *Computers in Human Behavior*.

Ninaus, M., Pereira, G., Stefitz, R., Prada, R., Paiva, A., Neuper, C., & Wood, G. (2015). Game elements improve performance in a working memory training task. *IJSG, 2*(1), 3-16.

Norman, D. (1993). Things That Make Us Smart, Defending Human Attributes in the Age of the Machine, p. 58.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 12*(5), 481-494.

O'Reilly, E. N. (2014). Correlations among Perceived Autonomy Support, Intrinsic Motivation, and Learning Outcomes in an Intensive Foreign Language Program. *Theory & Practice in Language Studies, 4*(7).

Olszewski, A. E., & Wolbrink, T. A. (2017). Serious gaming in medical education: a proposed structured framework for game development. *Simulation in Healthcare, 12*(4), 240-253.

Oppl, S., Reisinger, F., Eckmaier, A., & Helm, C. (2017). A flexible online platform for computerized adaptive testing. *International Journal of Educational Technology in Higher Education, 14*(1), 2.

Orion, N., & Hofstein, A. (1994). Factors that influence learning during a scientific field trip in a natural environment. *Journal of research in science teaching, 31*(10), 1097-1119.

Page, E. B. (1994). Computer grading of student prose, using modern concepts and software. *The journal of experimental education, 62*(2), 127-142.

Palaigeorgiou, G., Malandrakis, G., & Tsolopani, C. (2017). *Learning with Drones: flying windows for classroom virtual field trips.* Paper presented at the Advanced Learning Technologies (ICALT), 2017 IEEE 17th International Conference on.

Paleontological Research Institution. Real earth systems Science. Retrieved from http://virtualfieldwork.org/Virtual_Fieldwork.html

Papanastasiou, G., Drigas, A., & Skianis, C. (2017). Serious Games in Preschool and Primary Education: Benefits And Impacts on Curriculum Course Syllabus. *International Journal of Emerging Technologies in Learning (iJET), 12*(01), 44-56.

Parchman, S. W., Ellis, J. A., Christinaz, D., & Vogel, M. (2000). An evaluation of three computer-based instructional strategies in basic electricity and electronics training. *Military Psychology, 12*(1), 73-87.

Pasquier, P., Mérat, S., Malgras, B., Petit, L., Queran, X., Bay, C., . . . Auroy, Y. (2016). A serious game for massive training and assessment of French soldiers involved in forward combat casualty care (3D-SC1): development and deployment. *JMIR Serious Games, 4*(1).

Patiar, A., Ma, E., Kensbock, S., & Cox, R. (2017). Hospitality Management Students' Expectation and Perception of a Virtual Field Trip Web Site: An Australian Case Study Using Importance–Performance Analysis. *Journal of Hospitality & Tourism Education, 29*(1), 1-12.

Patiar, A., Ma, E., Kensbock, S., & Cox, R. (2017). Students' perceptions of quality and satisfaction with virtual field trips of hotels. *Journal of Hospitality and Tourism Management, 31*, 134-141.

Pellas, N., & Mystakidis, S. (2020). A systematic review of research about game-based learning in virtual worlds. *J. Univers. Comput. Sci*.

Perkins, C., Evans, M., Gavin, H., Johns, J., & Moore, J. (2001). Fieldwork and problem-based learning. *Planet, 4*(1), 27-28.

Pesare, E., Roselli, T., Corriero, N., & Rossano, V. (2016). Game-based learning and gamification to promote engagement and motivation in medical learning contexts. *Smart Learning Environments, 3*(1), 5.

Petri, G., von Wangenheim, C. G., & Borgatto, A. F. (2016). MEEGA+: an evolution of a model for the evaluation of educational games. *INCoD/GQS, 3*.

Peuquet, D. J. (1988). Representations of geographic space: toward a conceptual synthesis. *Annals of the Association of American Geographers, 78*(3), 375-394.

Pham, H. C., Dao, N., Pedro, A., Le, Q. T., Hussain, R., Cho, S., & Park, C. (2018). Virtual field trip for mobile construction safety education using 360-degree panoramic virtual reality. *Int. J. Eng. Educ, 34*(4), 1174-1191.

Piech, C., Sahami, M., Huang, J., & Guibas, L. (2015). *Autonomously generating hints by inferring problem solving policies.* Paper presented at the Proceedings of the Second (2015) ACM Conference on Learning@ Scale.

Pivec, M., Dziabenko, O., & Schinnerl, I. (2004). Game-based learning in universities and lifelong learning:"UniGame: social skills and knowledge training" game concept. *Journal of Universal Computer Science, 10*(1), 14-26.

Plant, R. W., & Ryan, R. M. (1985). Intrinsic motivation and the effects of self-consciousness, self-awareness, and ego-involvement: An investigation of internally controlling styles. *Journal of personality, 53*(3), 435-449.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE), 1*(1), 21-21.

Price, T. W., Dong, Y., & Barnes, T. (2016). *Generating data-driven hints for open-ended programming.* Paper presented at the Proceedings of the 9th International Conference on Educational Data Mining, International Educational Data Mining Society.

Procci, K., Singer, A. R., Levy, K. R., & Bowers, C. (2012). Measuring the flow experience of gamers: An evaluation of the DFS-2. *Computers in Human Behavior, 28*(6), 2306-2312.

Rahman, K. A., Ahmad, S., & Nordin, M. J. (2007). *The design of an automated c programming assessment using pseudo-code comparison technique.* Paper presented at the National Conference on Software Engineering and Computer Systems.

Ramaprasad, A. (1983). On the definition of feedback. *Systems Research and Behavioral Science, 28*(1), 4-13.

Roll, I., Aleven, V., McLaren, B. M., & Koedinger, K. R. (2011). Improving students' help-seeking skills using metacognitive feedback in an intelligent tutoring system. *Learning and Instruction, 21*(2), 267-280.

Roungas, B., & Dalpiaz, F. (2015). *A model-driven framework for educational game design.* Paper presented at the International Conference on Games and Learning Alliance.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist, 55*(1), 68.

Ryan, R. M., & Powelson, C. L. (1991). Autonomy and relatedness as fundamental to motivation and education. *The journal of experimental education, 60*(1), 49-66.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and emotion, 30*(4), 344-360.

Ryle, G. (1949). *The concept of mind*. Oxford, England: Barnes & Noble.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional science, 18*(2), 119-144.

Saikkonen, R., Malmi, L., & Korhonen, A. (2001). *Fully automatic assessment of programming exercises.* Paper presented at the ACM Sigcse Bulletin.

Sajjadi, P., Bagher, M. M., Cui, Z., Myrick, J., Swim, J. K., White, T. S., & Klippel, A. (2020a). *Cz investigator: Learning about critical zones through a vr serious game.* Paper presented at the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW).

Sajjadi, P., Bagher, M. M., Cui, Z., Myrick, J. G., Swim, J. K., White, T. S., & Klippel, A. (2020b). *Design of a Serious Game to Inform the Public About the Critical Zone.* Paper presented at the 2020 IEEE 8th International Conference on Serious Games and Applications for Health (SeGAH).

Sánchez, J., & Olivares, R. (2011). Problem solving and collaboration using mobile serious games. *Computers & Education, 57*(3), 1943-1952.

Santos, P., PéRez-SanagustíN, M., HernáNdez-Leo, D., & Blat, J. (2011). QuesTInSitu: From tests to routes for assessment in situ activities. *Computers & Education, 57*(4), 2517-2534.

Sauter, V. L. (1999). Intuitive decision-making. *Communications of the ACM, 42*(6), 109-115.

Scarce, R. (1997). Field trips as short-term experiential education. *Teaching Sociology, 25*(3), 219-226.

Schirm, J., Tullius, G., & Habgood, J. (2019). *Towards an objective measure of presence: examining startle reflexes in a commercial virtual reality game.* Paper presented at the Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts.

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments, 10*(3), 266-281.

Schuemie, M., Abel, B., van der Mast, C., Krijn, M., & Emmelkamp, P. (2005). *The effect of locomotion technique on presence, fear and usability in a virtual environment.* Paper presented at the EUROMEDIA.

Schultz, J. M. (2000). Computers and collaborative writing in the foreign language curriculum. *Network-based language teaching: Concepts and practice*, 121-150.

Scott, G. W., Goulder, R., Wheeler, P., Scott, L. J., Tobin, M. L., & Marsham, S. (2012). The value of fieldwork in life and environmental sciences in the context of higher education: A case study in learning about biodiversity. *Journal of Science Education and Technology, 21*(1), 11-21.

Scurati, G. W., Bertoni, M., Graziosi, S., & Ferrise, F. (2021). Exploring the use of virtual reality to support environmentally sustainable behavior: A framework to design experiences. *Sustainability, 13*(2), 943.

Seifan, M., Dada, O. D., & Berenjian, A. (2020). The effect of real and virtual construction field trips on students' perception and career aspiration. *Sustainability, 12*(3), 1200.

The Shadow of the Tomb Raider (Standard Edition) [Video game]. (2018). Canada: Eidos-Montréal.

Shena, B. S. A., Sitohang, B., & Rukmono, S. A. (2019). *Application of Dynamic Difficulty Adjustment on Evidence-centered Design Framework for Game Based Learning.* Paper presented at the 2019 International Conference on Data and Software Engineering (ICoDSE).

Shi, Y. R., & Shih, J. L. (2015). Game factors and game-based learning design model. *International Journal of Computer Games Technology, 2015*.

Shinneman, A. L., Loeffler, S., & Myrbo, A. E. (2020). Self-guided field trips allow flexibility in undergraduate student introductory field experiences. *Journal of Geoscience Education, 68*(4), 371-379.

Shute, V., Ke, F., & Wang, L. (2017). Assessment and adaptation in games. In *Instructional Techniques to Facilitate Learning and Motivation of Serious Games* (pp. 59-78): Springer.

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research, 78*(1), 153-189.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. *Computer games and instruction, 55*(2), 503-524.

Shute, V. J., & Ke, F. (2012). Games, learning, and assessment. In *Assessment in game-based learning* (pp. 43-58): Springer.

Shute, V. J., Leighton, J. P., Jang, E. E., & Chu, M. W. (2016). Advances in the science of assessment. *Educational Assessment, 21*(1), 34-59.

Shute, V. J., Ventura, M., Bauer, M., & Zapata-Rivera, D. (2009). Melding the power of serious games and embedded assessment to monitor and foster learning. *Serious games: Mechanisms and effects, 2*, 295-321.

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review, 20*(1), 21-53.

Singh, R., Gulwani, S., & Solar-Lezama, A. (2013). Automated feedback generation for introductory programming assignments. *ACM SIGPLAN Notices, 48*(6), 15-26.

Skarbez, R., Brooks, J., Frederick P, & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys (CSUR), 50*(6), 1-39.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 3*(2), 130-144.

Sofield, T. H., & Marafa, L. M. (2019). Revitalizing fieldtrips in tourism: Visual anthropology, photo elicitation, rapid appraisal, participant observation and Habermas. *Tourism Management, 75*, 522-546.

Sommerauer, P., & Müller, O. (2014). Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers & Education, 79*, 59-68.

Spacco, J., Hovemeyer, D., Pugh, W., Emad, F., Hollingsworth, J. K., & Padua-Perez, N. (2006). Experiences with marmoset: designing and using an advanced submission and testing system for programming courses. *ACM Sigcse Bulletin, 38*(3), 13-17.

Spires, H. A., Rowe, J. P., Mott, B. W., & Lester, J. C. (2011). Problem solving and game-based learning: Effects of middle grade students' hypothesis testing strategies on learning outcomes. *Journal of Educational Computing Research, 44*(4), 453-472.

Stephens, A. L., Pallant, A., & McIntyre, C. (2016). Telepresence-enabled remote fieldwork: Undergraduate research in the deep sea. *International Journal of Science Education, 38*(13), 2096-2113.

Stevens, J. A., & Kincaid, J. P. (2015). The relationship between presence and performance in virtual simulation training. *Open Journal of Modelling and Simulation, 3*(02), 41.

Stokes, A., Collins, T., Maskall, J., Lea, J., Lunt, P., & Davies, S. (2012). Enabling remote access to fieldwork: Gaining insight into the pedagogic effectiveness of ‘direct’and ‘remote’field activities. *Journal of Geography in Higher Education, 36*(2), 197-222.

Stott, T., Nuttall, A. M., & McCloskey, J. (2009). Design, development and student evaluation of a Virtual Alps Field Guide. *Planet, 22*(1), 64-71.

Streule, M., & Craig, L. (2016). Social learning theories—An important design consideration for geoscience fieldwork. *Journal of Geoscience Education, 64*(2), 101-107.

Subnautica. (Standard Edition)[Video game]. (2018). USA: Unknown Worlds Entertainment.

Sung, H. Y., & Hwang, G. J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education, 63*, 43-51.

Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*.

Thomas, G. J., & Munge, B. (2017). Innovative outdoor fieldwork pedagogies in the higher education sector: Optimising the use of technology. *Journal of Outdoor and Environmental Education, 20*(1), 7.

Triantafillou, E., Georgiadou, E., & Economides, A. A. (2008). The design and evaluation of a computerized adaptive test on mobile devices. *Computers & Education, 50*(4), 1319-1330.

Tsai, F. H., Tsai, C. C., & Lin, K. Y. (2015). The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. *Computers & Education, 81*, 259-269.

Tsigilis, N., & Theodosiou, A. (2003). Temporal stability of the intrinsic motivation inventory. *Perceptual and motor skills, 97*(1), 271-280.

Tüzün, H., Yılmaz-Soylu, M., Karakuş, T., İnal, Y., & Kızılkaya, G. (2009). The effects of computer games on primary school students' achievement and motivation in geography learning. *Computers & Education, 52*(1), 68-77.

University of Southampton. Virtual Fieldwork Project. Retrieved from http://visualisation.soton.ac.uk/

Ushaw, G., Davison, R., Eyre, J., & Morgan, G. (2015). *Adopting best practices from the games industry in development of serious games for health.* Paper presented at the Proceedings of the 5th International Conference on Digital Health 2015.

Ushaw, G., Eyre, J., & Morgan, G. (2017). *A paradigm for the development of serious games for health as benefit delivery systems.* Paper presented at the 2017 IEEE 5th International Conference on Serious Games and Applications for Health (SeGAH).

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education, 2*, 319-330.

Van der Pol, J., Van den Berg, B., Admiraal, W. F., & Simons, P. R. J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education, 51*(4), 1804-1817.

Van Eck, R. (2006). Digital game-based learning: It's not just the digital natives who are restless. *Educause Review, 41*(2), 16.

Vandercruysse, S., Vandewaetere, M., Cornillie, F., & Clarebout, G. (2013). Competition and students' perceptions in a game-based language learning environment. *Educational Technology Research and Development, 61*(6), 927-950.

Verdes, A., Navarro, C., & Álvarez-Campos, P. (2021). Mobile learning applications to improve invertebrate zoology online teaching. *Invertebrate Biology, 140*(1), e12321.

Verschueren, S., Buffel, C., & Vander Stichele, G. (2019). Developing theory-driven, evidence-based serious games for health: framework based on research community insights. *JMIR Serious Games, 7*(2), e11565.

Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational technology & society, 8*(2), 54-65.

Virvou, M., Manos, C., Katsionis, G., & Tourtoglou, K. (2002). *VR-ENGAGE: A virtual reality educational game that incorporates intelligence.* Paper presented at the Proceedings of IEEE international conference on advanced learning technologies.

Wang, L. C., & Chen, M. P. (2010). The effects of game strategy and preference-matching on flow experience and programming performance in game-based learning. *Innovations in Education and Teaching International, 47*(1), 39-52.

Wang, T. H. (2008). Web-based quiz-game-like formative assessment: Development and evaluation. *Computers & Education, 51*(3), 1247-1263.

Wang, Z., Gong, S. Y., Xu, S., & Hu, X. E. (2019). Elaborated feedback and learning: Examining cognitive and motivational influences. *Computers & Education, 136*, 130-140.

Ware, P. D., & Warschauer, M. (2006). Electronic feedback and second language writing. *Feedback in second language writing: Contexts and issues*, 105-122.

Weiss, D. J. (1974). Strategies of Adaptive Ability Measurement.

Weiss, D. J., & Kingsbury, G. (1984). Application of computerized adaptive testing to educational problems. *Journal of Educational Measurement, 21*(4), 361-375.

Welsh, K., & France, D. (2012). Spotlight on... Smartphones and fieldwork. *Geography, 97*, 47.

Welsh, K., Mauchline, A. L., Powell, V., France, D., Park, J. R., & Whalley, W. B. (2015). Student perceptions of iPads as mobile learning devices for fieldwork. *Journal of Geography in Higher Education, 39*(3), 450-469.

Wenger, E. (1998). Communities of practice: Learning as a social system. *Systems thinker, 9*(5), 2-3.

Westera, W., Nadolski, R., Hummel, H. G., & Wopereis, I. G. (2008). Serious games for higher education: a framework for reducing design complexity. *Journal of Computer Assisted Learning, 24*(5), 420-432.

Winarni, E. W., & Purwandari, E. P. (2018). Disaster Risk Reduction for Earthquake Using Mobile Learning Application to Improve the Students Understanding in Elementary School. *Mediterranean Journal of Social Sciences, 9*(2), 205-205.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225-240.

Wrzesien, M., & Raya, M. A. (2010). Learning in serious virtual worlds: Evaluation of learning effectiveness and appeal to students in the E-Junior project. *Computers & Education, 55*(1), 178-187.

Wu, P. H., Hwang, G. J., Milrad, M., Ke, H. R., & Huang, Y. M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British journal of educational technology, 43*(2), 217-232.

Xiao, A., Zeng, S., Allen, J. K., Rosen, D. W., & Mistree, F. (2005). Collaborative multidisciplinary decision making using game theory and design capability indices. *Research in Engineering Design, 16*(1-2), 57-72.

Yang, F., Li, F. W., & Lau, R. W. (2013). A fine-grained outcome-based learning path model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(2), 235-245.

Yang, J. C., Chien, K. H., & Liu, T. C. (2012). A digital game-based learning system for energy education: An energy conservation pet. *TOJET: The Turkish Online Journal of Educational Technology, 11*(2).

Yang, J. C., & Quadir, B. (2018). Individual differences in an English learning achievement system: Gaming flow experience, gender differences and learning motivation. *Technology, Pedagogy and Education, 27*(3), 351-366.

Yusoff, A., Crowder, R., Gilbert, L., & Wills, G. (2009). *A conceptual framework for serious games.* Paper presented at the Advanced Learning Technologies, 2009. ICALT 2009. Ninth IEEE International Conference on.

Zhao, J., LaFemina, P., Carr, J., Sajjadi, P., Wallgrün, J. O., & Klippel, A. (2020). *Learning in the field: Comparison of desktop, immersive virtual reality, and actual field trips for place-based STEM education.* Paper presented at the 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).

Zielke, M. A., Evans, M. J., Dufour, F., Christopher, T. V., Donahue, J. K., Johnson, P., . . . Flores, R. (2009). Serious games for immersive cultural training: Creating a living world. *IEEE computer graphics and applications, 29*(2), 49-60.

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38*(9), 25-32.