Journal of Computer Assisted Learning

WILEY

Journal of Computer Assisted Learning

ORIGINAL ARTICLE OPEN ACCESS

From Blocks to Artefacts: Exploring the Potential of Minecraft Education Edition in Archaeological Pedagogy

Joaquín Jiménez-Puerto | Gianni Gallello

Department of Prehistory, Archaeology and Ancient History (PREMEDOC Research Group), University of Valencia, Valencia, Spain

Correspondence: Joaquín Jiménez-Puerto (joaquin.jimenez@uv.es)

Received: 10 October 2024 | **Revised:** 11 February 2025 | **Accepted:** 17 April 2025

Keywords: archaeological stratigraphy | archaeology education | digital game-based learning | educational technology | experiential learning | minecraft education edition

ABSTRACT

Background: In the dynamic landscape of higher education, innovation through gamification has emerged as a successful approach to enhance student learning and motivation. Archaeological studies have particularly benefited from innovative pedagogy, as traditional teaching methods often struggle to engage younger generations.

Aims: To evaluate the effectiveness of Minecraft Education Edition (MEE) in teaching archaeological stratigraphy principles to History students, particularly when traditional fieldwork is not feasible.

Sample: Third-year History students ($n = 74$) at the University of Valencia (Spain), participating in the "Recent Universal Prehistory" course.

Method: A mixed-methods approach implementing both traditional teaching and MEE-based instruction. Student learning was assessed through pre-test, inter-test, and post-test evaluations, complemented by qualitative feedback analysis.

Results: Implementation of MEE demonstrated significant improvement in students' understanding of stratigraphic principles. The integration of technology and immersive experiences effectively bridged the gap between theory and practise, providing an interactive and experiential learning environment.

Conclusions: MEE proves to be an effective educational tool for teaching archaeological stratigraphy, particularly valuable when traditional fieldwork is not feasible. This innovative approach revitalises the teaching of crucial archaeological skills while providing a more engaging learning experience.

1 | Introduction

Archaeology, as a gateway to unravelling humanity's past, has undergone significant transformations in modern educational contexts, particularly in university settings. The teaching of complex concepts like archaeological stratigraphy often faces challenges in conveying information in an engaging and comprehensible manner. To address this, educators are increasingly turning to gamification, a pedagogical approach that incorporates game-design elements into non-game contexts, showing promise in enhancing student engagement and deepening understanding of archaeological methods (Deterding et al. 2011; Hamari et al. 2014).

Archaeological stratigraphy is the discipline of deciphering chronological and spatial relationships between layers of archaeological deposits and forms the backbone of archaeological excavation (Harris 1979b). Traditionally taught through lectures, textbooks, and two-dimensional diagrams, the integration of Minecraft Education Edition (MEE) (Mojang Studios 2024) into the curriculum represents an innovative shift. This approach not only makes learning about stratigraphy more interactive but also aligns with the needs of today's digital-native students (Prensky 2001).

The use of MEE has expanded beyond primary education to include secondary schools and universities. Numerous studies

Summary

- What is already known about this topic
  - Traditional methods of teaching archaeological stratigraphy often struggle to engage students effectively with complex spatial concepts.
  - The destructive nature of archaeological fieldwork limits hands-on learning opportunities for students.
  - Digital game-based learning has shown promise in enhancing student engagement and understanding in various educational contexts.
  - Minecraft Education Edition (MEE) has been successfully used as an educational tool in other disciplines.
- What this paper adds
  - Demonstrates the effectiveness of MEE in teaching archaeological stratigraphy, showing significant improvements in student understanding and retention compared to traditional methods.
  - Promotes heritage preservation by providing a non-destructive digital environment for novice archaeologists to practise excavation techniques.
  - Implements a robust mixed-methods approach combining quantitative analysis with qualitative feedback to evaluate the effectiveness of MEE in archaeological education.
  - Provides quantitative evidence of enhanced student performance, with mean test scores increasing from 5.64 (pre-test) to 9.04 (post-test) after MEE intervention.
  - Offers insights into how MEE enhances student engagement and practical application of complex spatial concepts in archaeology without risking actual archaeological sites.
  - Presents a model for integrating digital tools in archaeology education to create an interactive and experiential learning environment.
- Implications for practise and/or policy
  - Educators should consider incorporating MEE and similar digital tools into archaeology curricula to enhance learning outcomes and student engagement.
  - A blended approach, combining traditional teaching methods with interactive digital tools like MEE, should be adopted for comprehensive archaeological education.
  - Institutions should invest in digital infrastructure and training to support the implementation of game-based learning tools in archaeology programmes.
  - Further research and development should focus on creating more specialised archaeological simulations within MEE to cover a broader range of archaeological concepts and techniques.
  - Future research should focus on developing standardised assessment tools and conducting large-scale validation studies alongside creating specialised archaeological simulations within MEE.

have highlighted the potential of MEE as a gamification tool (AlJanah et al. 2023; Andersen and Rustad 2022; Callaghan 2016; Nkadimeng and Ankiewicz 2022). This versatile platform allows students to construct and explore archaeological sites in a virtual world, bridging theoretical knowledge with practical application. Our aim is to showcase how such innovative tools can transform educational practises, particularly in disciplines like archaeology that benefit immensely from visual and experiential learning (Kolb 2014).

In the teaching-learning process, both in school and university environments, it is essential to have tools that reinforce logical learning and creative thinking. These tools allow for the development of skills that enhance the resolution of complex problems. In this context, gamification methodologies help implement certain abilities (Gee 2003; Laurillard 2013; Mayer 2024). Digital technologies are significantly transforming society, the economy, and work dynamics. Alongside facing global challenges like climate change and the spread of misinformation, education systems must equip students for this evolving technological landscape (OECD 2019).

The integration of technology in education has shown promising results across various disciplines (Liu et al. 2016), making it a natural avenue for revitalising archaeological education. Traditional approaches to teaching archaeological stratigraphy have often relied on static images, diagrams, and lectures (Harris 1979b). While these methods provide a theoretical foundation, they can fall short in conveying the dynamic and three-dimensional nature of archaeological excavation. This deficiency poses a significant challenge for students transitioning from the classroom to fieldwork, where practical application of stratigraphic principles is paramount. Recent studies (Chenrai 2021; Gallello and Machause 2024; Garstki et al. 2019; Nixon and Price 2004; Waagen 2019) have emphasised the importance of adopting innovative pedagogical strategies to address this gap in geoscience and history education.

The use of educational games in archaeology has evolved significantly in recent years, with various platforms being employed to enhance student learning (Schrier 2018). Notable examples include the use of Assassin's Creed for teaching architectural history (Politopoulos et al. 2019), virtual reality applications for site reconstruction (Ellenberger 2017), and serious games specifically designed for archaeological education (Champion 2016). Recent studies have explored how digital games can support archaeological pedagogy through experiential learning (Morgan 2019), virtual fieldwork, and interactive site exploration (Boom et al. 2020; Fuchs et al. 2014).

However, while these applications have shown promise in various aspects of archaeological education, few have specifically addressed the teaching of stratigraphic principles. Traditional approaches to teaching stratigraphy have relied heavily on physical models and 2D representations, which often fail to convey the complex spatial relationships inherent in archaeological deposits. Recent attempts to use augmented reality for teaching stratigraphy have shown potential but face limitations in terms of accessibility and scalability. In this context, MEE offers a unique combination of accessibility, three-dimensional manipulation, and collaborative features that make it particularly suitable for teaching stratigraphic concepts.

MEE stands as a prominent digital game-based learning (DGBL) platform specifically crafted for educational use, developed from the original Minecraft software in 2011. With nearly 240 million units of the original Minecraft sold since its inception, MEE has emerged as a pioneering platform for educational innovation. By offering an interactive and customizable environment, MEE serves as an exceptional tool, specially developed for learning purposes (All et al. 2016). As of now, MEE has been used by over 35 million students and teachers from 115 countries since its release (Microsoft 2024).

As a sandbox game, MEE does not impose predefined objectives or goals, offering educators and students extensive creative freedom to shape their educational experiences. In this expansive virtual environment, students navigate through endlessly generative 3D worlds composed of cubic blocks, facilitating an interactive learning space that promotes exploration and creativity (Schrier 2018). MEE's classroom utility is further enhanced by its versatile features that support collaborative projects within the same virtual space, enable educators to monitor and adjust game settings to optimise learning, and provide tools for tracking progress and delivering targeted feedback.

Historically viewed by some as mere 'edutainment', recent perspectives validate the efficacy of DGBL in fostering meaningful educational experiences when integrated appropriately (Charsky 2010). This adaptability is supported by foundational educational theories such as Papert's constructionism, which suggests that learning is most effective when individuals engage in constructing tangible artefacts that facilitate deeper cognitive processing and discussion (Papert 1980).

Furthermore, MEE provides an ideal platform for project-based learning, a method characterised by student-centred inquiry and real-world problem solving that is deeply rooted in constructivist principles (Kokotsaki et al. 2016). Through MEE, educators can effectively implement project-based initiatives that foster transformative competencies and prepare students for future challenges by integrating critical thinking, collaboration, and digital literacy into their learning processes.

The power of MEE lies in its ability to simulate real-world phenomena within a virtual space, where the constraints of physical and financial resources are lifted. This virtual platform allows educators to create detailed reconstructions of archaeological sites, enabling students to explore and manipulate stratigraphic layers in a way that traditional classroom settings could never offer (Slattery et al. 2023). By recreating stratigraphic layers, students can actively participate in the process, gaining practical experience that transcends traditional classroom instruction.

Central to the educational value of MEE is its support for experiential learning. Students can engage directly with the material by building and excavating their own virtual archaeological sites. This hands-on approach helps to internalise the concepts of stratigraphy, such as understanding the chronological sequence of layers and interpreting historical artefacts found within those layers (Kolb 2014). The interactive nature of MEE not only deepens understanding but also increases student motivation. The familiar and enjoyable gameplay mechanics of MEE make learning more appealing and less intimidating, encouraging continuous engagement and exploration.

MEE offers a unique opportunity for teaching archaeology, particularly given the inherently destructive nature of the discipline. Traditionally, archaeology involves excavating sites that, once altered or excavated, cannot be restored to their original state, meaning that each dig can result in the irreversible loss of historical data. In this context, MEE provides a virtual solution that allows students to explore and experiment with archaeological techniques without the risk of damaging actual artefacts (Herrera and Riggs 2013; Kreager et al. 2023).

In conclusion, MEE transforms the teaching of archaeological stratigraphy from a passive learning experience into an active exploration. It offers a revolutionary approach that not only makes learning more interactive and enjoyable but also provides a deep, meaningful understanding of complex archaeological principles. As educators and institutions continue to seek innovative teaching methods, MEE stands out as a powerful tool that can reshape educational landscapes and significantly enhance learning outcomes.

1.1 | Objectives

The primary objective of this study is to compare the effectiveness of traditional teaching methods against the use of MEE in imparting the fundamental principles of archaeological stratigraphy. Specifically, the study aims to evaluate how MEE's interactive learning environment transforms the learning experience by allowing students not just to receive information, but to interact with it, thus enhancing their conceptual understanding of stratigraphy. This includes assessing MEE's ability to foster teamwork through its collaborative nature, and its impact on developing logical thinking and problem-solving skills by challenging students with tasks that require creative and logical solutions. As we navigate the evolving landscape of education, it becomes evident that the integration of technology, experiential learning, and virtual reconstruction offers a revolutionary way to bridge the gap between theory and practise in archaeology. This study seeks to contribute to this growing body of knowledge and provide insights into the effective implementation of digital game-based learning in archaeological education.

2 | Material and Method

The module "Recent Universal Prehistory" (code 34022) from the second year History Degree at the University of Valencia focuses on an integral study of humanity from the appearance of agriculture to the development of writing, encompassing significant economic, social, political, and cultural changes in Europe and the Near East during this period. A crucial component of the course is archaeological stratigraphy, a competency that students develop and refine through the analysis of archaeological and paleoclimatic data, enabling them to interpret environmental changes and their impacts on human societies.

The "Recent Universal Prehistory" module is designed with a dual focus on comprehensive theoretical knowledge and practical skills application, aimed at equipping students for professional roles in archaeology. It covers the expansive timeline from the Neolithic period to the Bronze Age, getting into the evolution

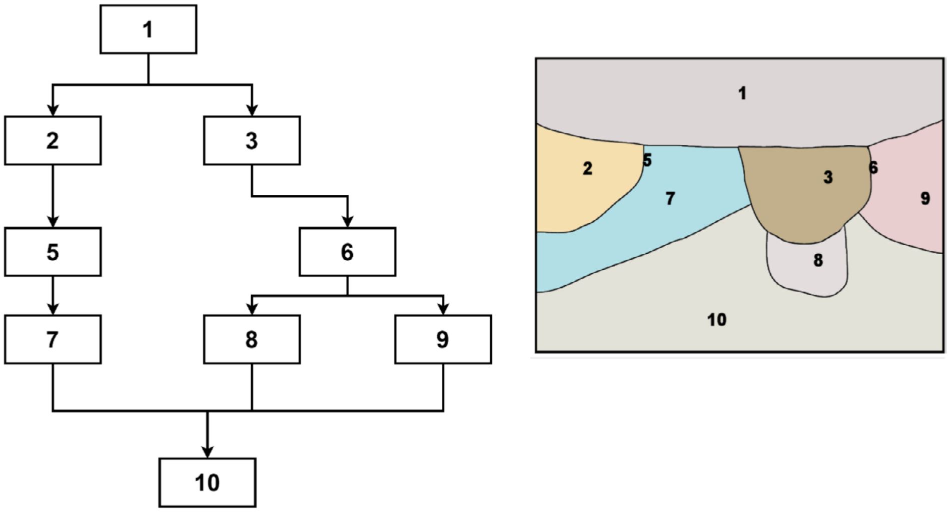

FIGURE 1 | Example of a section with the associated Harris Matrix.

of human societies from nomadic hunter-gatherers to settled agricultural communities and the ensuing significant shifts in social structures, economies, and cultures. A key emphasis is placed on developing the students' ability to analyse archaeological and paleoclimatic data to interpret the impact of environmental changes on historical human societies. Additionally, the course explores the first major technological innovations like agriculture and metallurgy, understanding their role in shaping complex social organisations and cultural expressions. The goal is to provide students with a robust foundation of knowledge and practical analytical skills necessary for digging deep into the layers of human history and contributing effectively to their future archaeological endeavours.

Integrating MEE into the "Recent Universal Prehistory" module aligns seamlessly with the course's core objectives. By utilising this innovative tool, the course can enhance its approach to teaching archaeological stratigraphy, providing a more interactive and engaging learning experience. MEE offers a virtual environment where students can simulate excavation processes and stratigraphic analysis, thereby deepening their understanding of human evolutionary stages and cultural transformations from the Neolithic to the Bronze Age. This method not only supports the course's aim to develop practical analytical skills by allowing students to visualise and manipulate data in a dynamic way but also strengthens their ability to apply theoretical knowledge to real-world archaeological scenarios. Thus, the use of MEE supports the overarching goal of the course to prepare students thoroughly for professional roles in archaeology, making the learning process both comprehensive and compelling.

The proposed method is based on Kolb's Experiential Learning Cycle (1984), which emphasises the importance of concrete experiences followed by reflective observation, abstract conceptualisation, and active experimentation. By actively engaging students in the excavation process within the virtual world of Minecraft, this approach facilitates a deeper understanding of stratigraphic principles. Utilising MEE as a platform for teaching archaeological stratigraphy involves digitally recreating archaeological sites. This process involves meticulous attention to detail, ensuring accuracy in the representation of stratigraphic layers, artefacts, and features.

Some specific methods are employed in Archaeology to reflect the reality of an excavation (Carver 2013; Renfrew and Bahn 2007). One of them is the Harris Matrix (Harris 1979a) which is a fundamental tool in archaeology used to organise and understand the chronological sequence of layers and structures uncovered during excavations. Think of an archaeological site as a layered cake, where each layer represents a different moment in time. As archaeologists excavate, they carefully remove these layers one by one, recording each discovery meticulously.

To create a Harris Matrix, archaeologists assign a unique number to each layer or feature found, known as a "context." They then construct a visual diagram where these contexts are represented as boxes or nodes. These boxes relate to lines that indicate which layers were found above or below others (Figure 1). This way, the Harris Matrix clearly shows the temporal relationship between different contexts, allowing archaeologists to understand what happened first, what came next, and how events unfolded over time at that site.

This method is essential for reconstructing the history of an archaeological site, as it helps to chronologically order the findings and interpret how human activities and natural changes occurred. The Harris Matrix, in essence, is a kind of time map that aids archaeologists in deciphering the history buried beneath the ground. Part of the proposed tasks for the students is to learn this methodology, which is not easy at the beginning. The Stratigraphic Units (SU) are not always easily recognised, and many times the interpretation is variable. It is precisely in this issue that MEE can provide new approaches and tools to Archaeology teachers.

The test was carried out in the group D of the module "Recent Universal Prehistory" (code 34022) from the second year History Degree at the University of Valencia. The MEE intervention consisted of six 2-h sessions spread across 3 weeks, ensuring sufficient exposure while maintaining engagement. This duration was chosen based on preliminary testing and

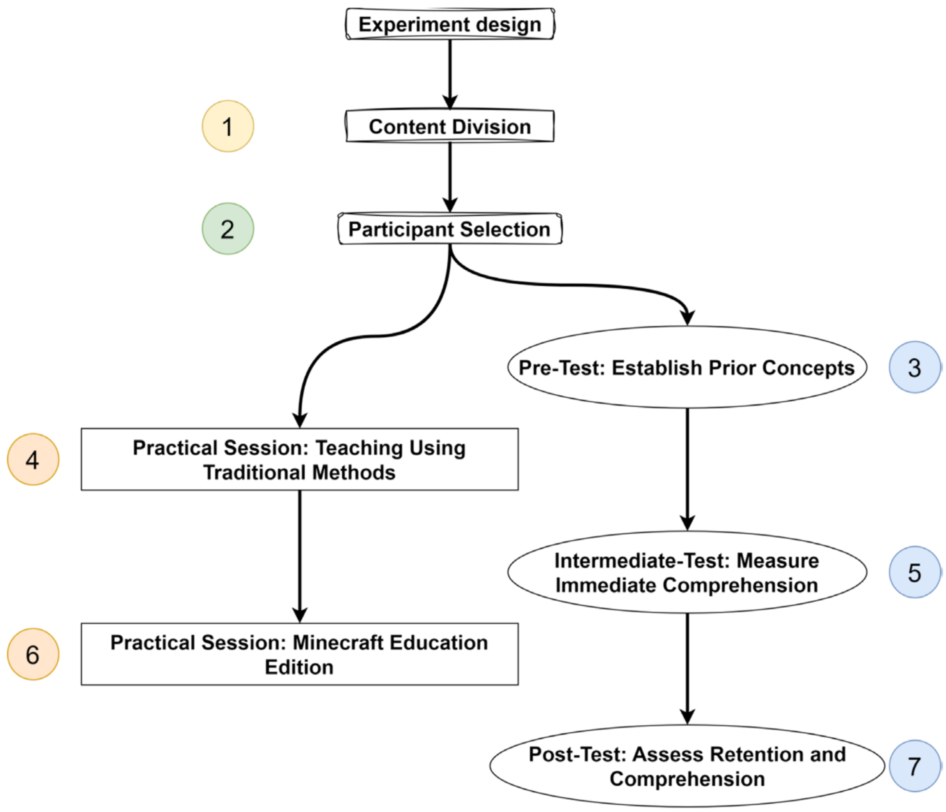

FIGURE 2 | Experiment design.

course scheduling constraints. In the experience 74 students divided in two subgroups were participating. Both were integrated by a diverse group with variations in age, gender, and social background, with the majority aged between 20 and 25 years. They also displayed varying levels of prior knowledge in archaeology. They were divided in two groups, one of 48 (Wednesdays mornings from 12:00 to 14:00 AM) and another of 26 (Tuesdays from 15:00 to 17:00 PM), from September to December 2023.

The design of the experiment is a critical phase that sets the foundation for evaluating the effectiveness of MEE in teaching archaeological stratigraphy compared to traditional methods. This chapter outlines the systematic approach adopted to structure and implement the experiment, ensuring that it meets rigorous academic standards and provides reliable results. The experiment is crafted to encompass various stages, from initial planning and content division to participant selection and practical sessions. Each step is designed to address specific research questions and objectives, allowing for a comprehensive analysis of the pedagogical impact of MEE. By integrating traditional teaching methods with innovative digital tools, this experiment aims to provide insights into how interactive and immersive environments can enhance learning outcomes in archaeology. Through this structured and methodical approach, the experiment seeks to shed light on the potential benefits and challenges of using digital game-based learning platforms like MEE in archaeological education. The findings from this experiment will contribute to the broader discourse on educational innovation and the integration of technology in teaching complex subjects such as archaeological stratigraphy.

The initial step in this process is the design of the experiment (Figure 2). The primary aim is to assess the effectiveness of MEE in enhancing students' understanding of archaeological stratigraphy compared to traditional teaching methods. Following the experiment design, the next step is the division of content. The curriculum is structured to cover essential concepts of archaeological stratigraphy. The content is divided into two main segments: one for traditional teaching methods and the other for MEE-based instruction. Each segment includes theoretical components, practical exercises, and assessments designed to measure comprehension and retention. The division ensures that both teaching methods cover the same material, allowing for a fair comparison of their effectiveness. Moreover, the content taught using both methods was carefully matched in terms of difficulty and relevance. This equivalence was crucial for validating the effectiveness of each instructional approach. Both teaching methods covered the same topics, allowing for a direct comparison of their impact on student learning outcomes. However, the difficulty of each successive test was slightly increased to account for the students' prior exposure to the concepts. This adjustment was necessary to prevent bias and accurately reflect the progressive learning process.

A preliminary assessment was conducted to evaluate the students' existing knowledge of stratigraphy before any teaching interventions. This test established a baseline for the research. Subsequently, an initial block of traditional instruction was delivered to familiarise students with fundamental concepts in archaeological stratigraphy. All conducted tests consisted of 10 multiple-choice questions. The maximum possible score for each test was 10 points, with each correct answer scoring

1 point. This standardised format allowed for consistent measurement of student knowledge across different stages of the study.

Following this initial set of classes, an intermediate test was administered to gauge the level of understanding achieved through conventional teaching methods. Next, the MEE (Multimedia Educational Environment) series of classes was implemented.

Finally, a post-test was conducted to assess the knowledge gained after the MEE intervention. This comprehensive approach allowed for a comparison of the students' initial knowledge, their understanding after traditional instruction, and their comprehension following the MEE intervention. By quantifying these differences, it became possible to evaluate the effectiveness of each teaching method. All the different tests and the proposed questions can be found in the Supporting Information section.

From a psychometric perspective, we recognise that formal test validation typically involves multiple components: content validity, construct validity, criterion-related validity, and reliability measures such as internal consistency (often measured through Cronbach's alpha) and test–retest reliability (Crocker and Algina 1986; Sireci and Padilla 2014). While comprehensive psychometric analysis would be valuable for standardising these assessment tools for wider use, our current focus was on ecological validity within the specific educational context. The assessment structure deliberately followed established cognitive frameworks, particularly Bloom's Taxonomy (Anderson and Krathwohl 2001), ensuring questions progressed from basic recall through application to analysis. The parallel construction of items across the three tests (pre-test, inter-test, and post-test) was designed to maintain consistent cognitive demands while varying specific content, a strategy aligned with contemporary psychometric principles (Bandalos 2018; Hambleton and Jones 1993; Price 2017). This approach, while not formally psychometrically validated, provided a pragmatic foundation for measuring learning gains in this specific educational context.

2.1 | Pre-Test: Establish Prior Concepts

Before the commencement of the practical sessions, all participants undergo a pre-test. This test is designed to establish their baseline knowledge of archaeological stratigraphy. It includes questions on key concepts, terminologies, and basic principles. The results of the pre-test help in understanding the participants' initial level of comprehension and in tailoring the subsequent teaching sessions to address any gaps in knowledge.

The development of the pre-test questions followed a rigorous methodological approach. A panel of expert archaeologists and educators collaboratively created a question bank, ensuring content validity. The questions were designed to assess various cognitive levels, from basic recall to application and analysis, following Bloom's Taxonomy (Anderson and Krathwohl 2001).

The pre-test content focuses on assessing students' baseline knowledge across several key areas of archaeological stratigraphy. It introduces fundamental terms and principles, including the science of analysing earth layers, the concept of stratigraphic units, and the importance of archaeological context. Questions cover essential principles of stratigraphy, such as superposition and original horizontality, and evaluate understanding of relative dating techniques used in archaeology. Additionally, the test examines students' grasp of the relevance and application of stratigraphy, particularly its importance in understanding historical contexts and sequences.

2.2 | Practical Session: Teaching Using Traditional Methods

Participants undergo a practical session that employs conventional teaching tools such as lectures, textbooks (Harris 1979b), academic articles (Kreager et al. 2023) and two-dimensional diagrams. This session aims to provide a thorough understanding of archaeological stratigraphy using tried-and-tested pedagogical approaches. Instructors focus on explaining the chronological and spatial relationships between archaeological layers and the significance of stratigraphic analysis in uncovering the history of a site.

Lectures were carefully prepared, covering key topics from historical principles to practical applications, complemented by visual presentations and Q&A sessions. Selected textbooks offer comprehensive coverage, serving both for pre-reading and in-class discussions. Custom-made two-dimensional diagrams include cross-section drawings, Harris matrices, and flowcharts illustrating stratigraphic interpretation processes.

The teaching methodology is based on sequential learning, with repetition and reinforcement of key concepts, balancing direct instruction with opportunities for active student engagement, and making extensive use of visual aids to support the comprehension of spatial concepts.

2.3 | Intermediate-Test: Measure Immediate Comprehension

The intermediate test serves as a crucial assessment point, designed to measure students' immediate comprehension following conventional practical sessions. Developed through a rigorous process involving both archaeological and pedagogical expertise, the test aligns with key module competencies including theoretical understanding of stratigraphic principles, practical application of analysis techniques, critical thinking in data interpretation, and proficiency in documentation.

The inter-test content builds upon the initial assessment, continuing to evaluate students' comprehension of key stratigraphic principles, including lateral continuity and cross-cutting relationships. It places a strong emphasis on interpretation and context, highlighting the crucial role of context in analysing archaeological findings and its influence on data interpretation. The test also introduces questions about traditional educational methods used in archaeology, assessing their effectiveness in teaching stratigraphy. Furthermore, it explores the challenges students face in visualising spatial and temporal relationships in stratigraphy.

From a pedagogical perspective, the intermediate test serves multiple purposes: it acts as a diagnostic tool, motivates active engagement with the material, encourages instructors to reflect on their teaching methods' effectiveness, and prepares students for more advanced concepts.

2.4 | Practical Session: Minecraft Education Edition

Participants engage in a practical session using MEE. This session leverages the immersive and interactive environment of Minecraft to teach archaeological stratigraphy. Instructors propose initiation exercises to familiarise students with the digital game-based learning (DGBL) environment and MEE interface. After ensuring everyone is comfortable with the game interface, the teachers recreate a detailed and invented archaeological site within the game, allowing students to explore and manipulate stratigraphic layers virtually. The hands-on nature of this session facilitates experiential learning, enabling students to actively engage with the material and gain practical experience in a simulated excavation environment.

2.5 | Post-Test: Assess Retention and Comprehension

The final phase of the experiment involves a post-test conducted after a predetermined period. This test assesses the long-term retention and comprehension of the material covered. It includes questions like the intermediate test, but with a slightly increased level of difficulty. The results of the post-test help in evaluating the sustained impact of both traditional and Minecraft-based teaching methods on students' knowledge and skills.

The post-test content focuses on evaluating the overall improvement in students' understanding and application of archaeological stratigraphy concepts. It assesses the application of stratigraphic principles using digital tools like Minecraft Education, highlighting the immersive learning experiences provided. The test emphasises the benefits of digital tools, particularly the interactive visualisation of stratigraphic layers and hands-on learning opportunities. Additionally, it includes a comparative analysis of the effectiveness of traditional and digital methods in improving understanding, with a particular emphasis on how technology can enhance learning outcomes. Finally, the post-test evaluates improvements in critical skills such as analysis and spatial reasoning, which are facilitated using digital simulations in the learning process.

This structure allows tracking the progression of students' understanding and skills throughout the course, from basic comprehension to practical application and critical analysis (Laurillard 2013; Lave and Wenger 1991; Piaget 1977).

2.6 | Question Mapping

The development and validation of the assessment instruments drew upon the authors' extensive experience teaching Archaeological Methodology at the University of Valencia. With over 15 years of combined experience in teaching archaeological stratigraphy concepts to undergraduate students, the authors have developed a deep understanding of common learning challenges and key assessment points in this field.

The question mapping strategy (detailed in Table 1) was designed to ensure comprehensive coverage of key stratigraphic principles while maintaining consistent difficulty levels across the three tests. Questions were carefully structured to assess both factual knowledge and conceptual understanding, reflecting the hierarchical nature of learning in archaeological stratigraphy. The assessment design was informed by years of observing student progression patterns and identifying critical points in the learning process of stratigraphic principles.

Special attention was paid to:

- Content validity: Questions directly aligned with core stratigraphic concepts taught in the course
- Progressive complexity: Tests maintained comparable overall difficulty while introducing increasingly sophisticated applications
- Practical relevance: Assessment items connected theoretical knowledge with practical field applications.

The strong effect sizes observed and the consistent patterns of improvement across different question types suggest the instruments were effective in capturing learning gains. While formal psychometric validation was beyond the scope of this initial proof-of-concept study, the assessment design was grounded in substantial practical experience with teaching and evaluating archaeological stratigraphy concepts.

To ensure valid comparisons across the pre-test, inter-test, and post-test, a systematic question mapping approach was implemented. This methodology allowed for a direct assessment of how students' understanding of specific topics evolved with different teaching methods and at various stages of the experiment. The questions were carefully aligned by theme throughout the three tests, focusing on five key areas of archaeological stratigraphy:

1. Principle of Superposition.
2. Relative Dating and Use of Stratigraphy.
3. Stratigraphic Unit and Archaeological Context.
4. Importance of Stratigraphy and Context.
5. Principle of Original Horizontality and Stratigraphic Correlation.

For each thematic area, corresponding questions were developed for the pre-test, inter-test, and post-test. These questions (from Q1 to Q10) were designed to progressively assess students' understanding, moving from basic concept recognition to more complex application and analysis. Table 1 presents an overview of the question mapping across the three tests.

This structured approach to question mapping ensured consistency in the assessment of key concepts across all three tests while allowing for an evaluation of the progression in students' understanding. It facilitated a comprehensive analysis of the effectiveness of both traditional and digital game-based learning methods in teaching archaeological stratigraphy.

2.7 | Qualitative Feedback

The last, but not least, a qualitative feedback field was set up for the students to comprehensively evaluate the entire process. This was done through an open-ended question that allowed participants to express their opinions about the whole experience in up to 300 characters. This approach enabled us to gather more nuanced and detailed insights into the students' perceptions, complementing the quantitative data collected throughout the study. The free-form responses provided valuable information about the effectiveness of the teaching methods, the engagement level of the activities, and any suggestions for improvement, offering a holistic view of the educational experience from the students' perspective.

2.8 | Crafting the "World"

MEE serves as an innovative platform for teaching archaeological stratigraphy through the digital recreation of archaeological sites. This process demands meticulous attention to detail, ensuring accurate representation of stratigraphic layers, artefacts, and features. The stratigraphy practise using MEE was structured in two phases, both featuring hypothetical cases based on typical archaeological scenarios. These digital reconstructions, while not replicating specific real-world sites, were carefully crafted to embody common stratigraphic challenges and features encountered in actual archaeological excavations, particularly from European prehistoric contexts. This focus aligns with the thematic scope of the broader educational module. By simulating these environments, the approach provides students with a realistic yet controlled learning experience, bridging the gap between theoretical knowledge and practical application in archaeological stratigraphy. The MEE crafted scenarios can be found in the Supporting Information.

In the first phase, three very simple and limited scenarios were designed (see Figure 3b,c) to help students become familiar with the game's interface and the basic concepts of archaeological excavation. The difficulty of these three initial exercises is progressive, featuring common situations such as horizontal stratigraphy and more complex habitational structures.

In the second phase, an "open world" excavation was proposed, situated in a cave (see Figure 3a). In all these experiences, students were later required to propose the corresponding Harris Matrix for the stratigraphic units (SU) found, following the stratigraphic principles introduced previously. This approach not only enhances students' understanding of archaeological

TABLE 1 | Question mapping across tests.

Thematic area	Pre-test	Inter-test	Post-test
1. Principle of superposition	Q2: Basic principle identification	Q2: Application to geological structures	Q3: Contextual application to artefact layers
2. Relative dating and use of stratigraphy	Q3: Identification of dating technique	Q3: Understanding of relative dating applications	Q1: Which aspect of stratigraphy helps determine the sequence
3. Stratigraphic unit and archaeological context	Q4: Definition of stratigraphic unit	Q4: Traditional teaching methods	Q4: Digital representation in Minecraft
4. Importance of stratigraphy and context	Q5: Definition of archaeological context	Q5: Interpretive importance of context	Q5: Critical skills in stratigraphic interpretation
5. Principle of original horizontality and stratigraphic correlation	Q6: Identification of temporal sequencing principle	Q6: Understanding of lateral continuity	Q7: Skill development through digital simulation

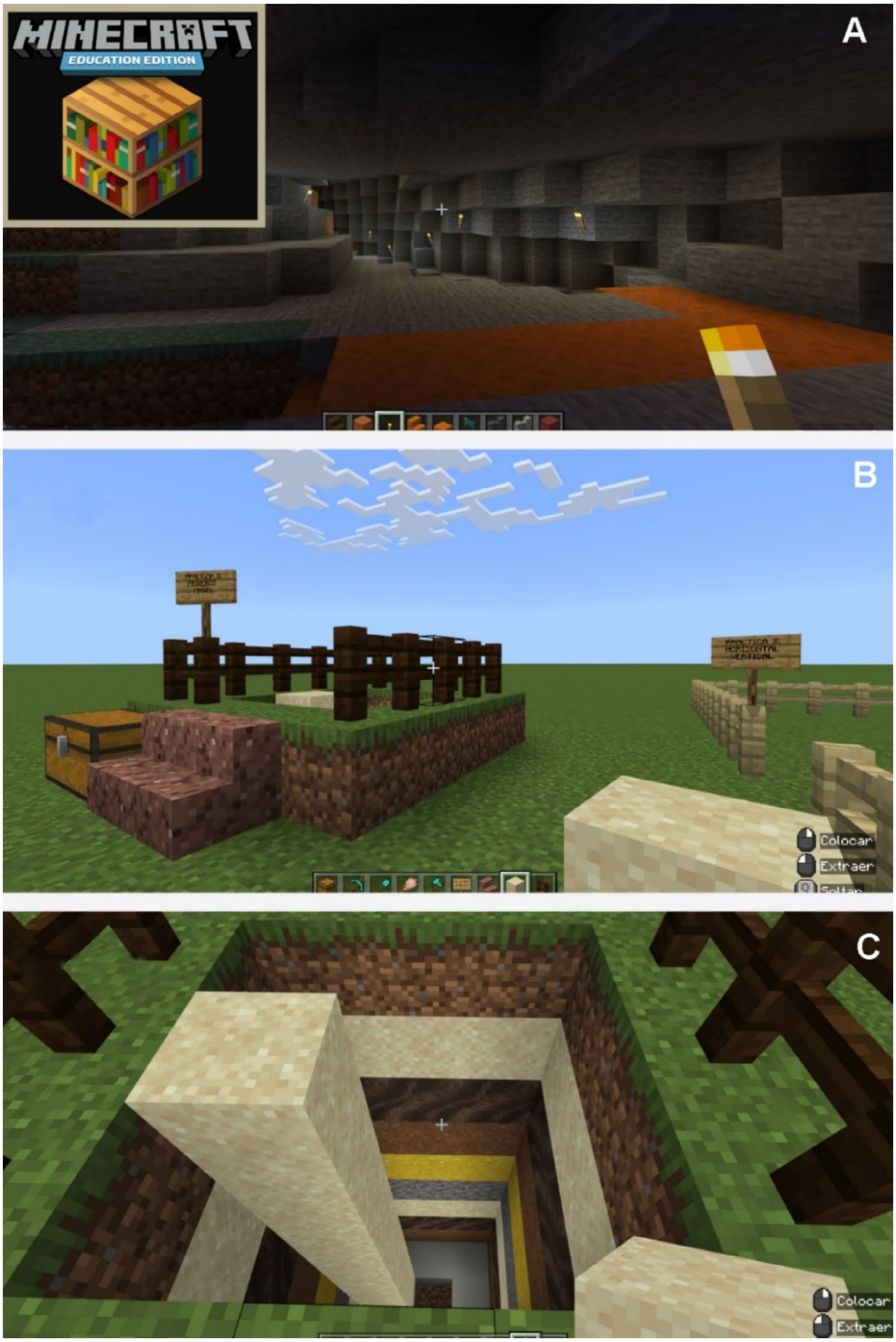

FIGURE 3 | Images of the designed MEE landscapes.

stratigraphy but also provides an interactive and engaging way to apply theoretical concepts in a practical, simulated environment.

The first phase, created ad hoc in a flat, controlled environment, consisted of three progressively challenging tests: a simple stratigraphic sequence with contrasting colours and no archaeological remains, a more complex sequence featuring basic archaeological structures, and an excavation with horizontal stratigraphy of pit fields. Each layer was meticulously designed using specific Minecraft colour blocks to represent different soil types and materials. The colour choices were based on the need to clearly contrast the layers, facilitating students' visual identification of stratigraphy. This approach allowed students to gradually familiarise themselves with the game interface and basic excavation concepts.

The second phase, with a more advanced scenario, recreated a cave excavation within a Minecraft world with natural relief. This environment was designed to simulate the complexities of a real archaeological site, including a secret chamber filled with sediment concealing a rock art mural, challenging students to interpret it correctly. Creating this scenario involved careful manipulation of Minecraft terrain, strategic placement of blocks to simulate geological strata, and incorporation of subtle archaeological elements. A cave environment was chosen for its relevance in many archaeological contexts and its potential to present complex stratigraphic challenges. In both scenarios, students were required to apply stratigraphic principles to create the corresponding Harris Matrix. The design process did not utilise specific algorithms but relied on manual construction of each layer and element, allowing precise control over the complexity and challenges presented at each level.

Additionally, in all cases, the practise could be conducted individually or with multiple students collaborating simultaneously in the same scenario, which enhanced the playful aspect of the experience. Due to the nature of the practise, students were also able to explore aspects that would be impossible to implement in the real world, including conducting the excavation carelessly. This allowed them to observe firsthand the effects of "bad practises" and the impact they had on the scenario. After some initial attempts where many got carried away by the excitement of the brand-new environment, all groups eventually carried out their work meticulously and correctly. This approach not only provided a fun and interactive learning experience but also emphasised the importance of careful work in archaeological excavations.

The interactive nature of Minecraft allows for hands-on learning, where students directly manipulate the virtual environment. Additionally, the multiplayer functionality fosters collaborative learning experiences, simulating the teamwork often required in archaeological fieldwork.

2.9 | Data Processing and Statistics

To evaluate the effectiveness of the educational intervention, a comprehensive statistical analysis was conducted on the students' performance across three test phases: Pre-test, inter-test, and post-test. The individual test scores for each student are available in the Supporting Information section. This multiphase testing approach is widely recognised in educational research for its ability to track changes in student knowledge and skills over time (Cohen et al. 2002). The analysis began with descriptive statistics, calculating the mean, median, mode, and standard deviation for each test to provide an overview of the score distributions. These measures offer a foundational understanding of the data and are crucial for identifying general trends and potential outliers (Field 2024).

To assess the overall impact of the intervention, a one-way Analysis of Variance (ANOVA) was performed to compare the means across all three test results, followed by Tukey's Honest Significant Difference (HSD) post hoc test to identify specific differences between test pairs. ANOVA is particularly useful in educational studies for comparing multiple groups simultaneously (Gelman 2005), while Tukey's HSD test provides a conservative approach to multiple comparisons, reducing the risk of errors (Abdi and Williams 2010).

Further in-depth analysis included paired t-tests between each pair of tests (pre-test vs. inter-test, inter-test vs. post-test, and pre-test vs. post-test) to determine statistically significant changes in student performance at each stage of the intervention. Paired t-tests are especially appropriate for this design as they account for the within-subject nature of the data (Zimmerman 1997). The magnitude of the intervention's effect was quantified using Cohen's d effect size calculation between the pre-test and post-test scores, providing a standardised measure of the overall impact. Effect size measures are increasingly emphasised in educational research as they provide information about the practical significance of results, complementing statistical significance (Sullivan and Feinn 2012). Additionally, an improvement analysis was conducted by calculating the mean difference between post-test and pre-test scores to quantify the average gain in student performance. This approach aligns with recommendations in the literature for assessing learning gains in educational interventions (Hake 1998).

To implement the question mapping approach, we conducted paired sample t-tests to compare student responses across the different test phases. We first analysed the differences between the pre-test and inter-test scores, followed by a comparison of the inter-test and post-test results. These statistical analyses allowed us to determine the statistical significance of changes in mean scores, providing an objective assessment of our educational interventions' effectiveness at each stage of the experiment. By employing this method, we were able to quantify the distinct impacts of traditional teaching methods and the integration of MEE on students' comprehension of archaeological stratigraphy. This comprehensive analytical strategy offered insights into the progressive development of student understanding throughout the course of the study, aligning with best practises in educational research methodology (Creswell and Creswell 2017).

To gain even deeper insights into the multidimensional nature of student learning in archaeological stratigraphy, we employed principal component analysis (PCA) on the test scores from all three phases (pre-test, inter-test, and post-test). This multivariate statistical technique was utilised to reduce the dimensionality of our data while preserving its essential patterns and relationships (Jolliffe 2002). By applying PCA, we aimed to identify the underlying structures in the students' responses, revealing how different concepts in archaeological stratigraphy are interrelated in the learning process. The analysis was performed on the question-level data, with each student's responses across all tests serving as individual observations. The PCA was conducted using a dataset comprised of responses from 74 students to 10 questions on archaeological stratigraphy. These questions were administered at three distinct phases of the study: pre-test, inter-test, and post-test, resulting in a total of 222 observations (74 students $\times$ 3 phases). Each observation encompasses 10 variables, corresponding to the responses to the 10 test questions. These variables capture the students' knowledge and understanding of key concepts in archaeological stratigraphy at different stages of their learning. The pre-test assessed students' initial knowledge, the inter-test measured their progress during the course, and the post-test evaluated their final understanding of the subject matter. This data structure allows us to analyse not only the differences between students but also the progression of learning over time for everyone. We focused on the first two principal components, which captured the most significant variability in the data, to visualise the progression of student understanding. To visualise the results, we created a scatter plot of these components, with each point representing a student's performance across all questions. To better illustrate the distribution and separation of the three test phases, we encircled each group (pre-test, inter-test, and post-test) with 95% confidence ellipses.

This approach allowed us to map the learning trajectories of students from their initial state (pre-test) through the traditional teaching phase (inter-test) to the final stage after using Minecraft: Education Edition (post-test). By examining the loadings of individual questions on these components, we sought to understand which aspects of archaeological stratigraphy knowledge were most influential in distinguishing student performance and how these aspects evolved throughout the learning process. This method of using PCA to analyse educational data has been successfully employed in previous studies to uncover patterns in student learning and to evaluate the effectiveness of educational interventions (Lee et al. 2008).

This clear and detailed structure not only facilitates understanding of the study design but also underscores the systematic methodology used to evaluate the effectiveness of teaching strategies, which is essential for the validity and reproducibility of the study.

3 | Results

The analysis employed both descriptive and inferential statistical methods to evaluate the effectiveness of the educational intervention. Descriptive statistics provided an initial overview of student performance trends, while inferential testing through one-way ANOVA, paired *t*-tests, and PCA enabled robust examination of the intervention's impact. The statistical approach was structured to address three key aspects of the intervention:

- Overall performance changes: Analysed through descriptive statistics and one-way ANOVA to identify significant differences across test phases.
- Specific learning gains: Evaluated using paired *t*-tests to compare performance between consecutive test phases.
- Multidimensional learning patterns: Examined through PCA to understand the underlying structure of student learning progress

This comprehensive statistical framework allowed for a thorough evaluation of both the magnitude and nature of improvements in student understanding of archaeological stratigraphy concepts. This chapter presents the findings of our study, comparing the effectiveness of traditional teaching methods versus MEE in teaching archaeological stratigraphy. The results were obtained from a detailed analysis of pre-test, inter-test, and post-test scores (see Supporting Information), as well as qualitative feedback from the students. Our objective is to provide a comprehensive understanding of how each method impacts student learning outcomes. Both pre-test and post-test were scored objectively. Each multiple-choice question had equal weight, allowing for straightforward comparison of scores across the tests. This consistent scoring method ensures that the evaluation of student performance is fair and unbiased.

The results are presented in the following sections with detailed tables and charts illustrating the comparison of scores between pre-test and post-test. Qualitative insights are also discussed to provide a comprehensive understanding of the impact of MEE on learning archaeological stratigraphy.

3.1 | Test Scoring and Statistics

Statistical analysis, including both descriptive statistics and inferential testing (one-way ANOVA, paired *t*-tests, and Principal Component Analysis), revealed significant trends in student performance. Mean scores showed a marked increase from pre-test to post-test, while median scores reflected an overall improvement in central tendency. The mode highlighted common achievement levels, and a notable reduction in standard deviation from pre-test to post-test underscored the increased consistency in student performance after utilising MEE.

This comprehensive statistical overview demonstrated that areas of study employing MEE exhibited more substantial enhancements in student comprehension compared to those taught through traditional methodologies. Such improvement can be attributed to MEE's interactive and engaging nature, which enables students to visualise and manipulate stratigraphic layers within a virtual environment, thereby reinforcing their grasp of these concepts (See Table 2). The collective analysis of these descriptive statistics provides robust evidence of MEE's effectiveness in improving both the overall performance and the consistency of student understanding in archaeological stratigraphy.

An intriguing pattern emerged in the score distribution across the three tests. The pre-test established a baseline with a mean of 5.64 and a standard deviation of 1.19. The inter-test, following traditional teaching methods, showed a modest improvement in the mean (6.05) but an increased standard deviation (1.29), suggesting more diverse outcomes among students.

The Post-test results, however, were striking. The mean score rose significantly to 9.04, while the standard deviation decreased to 0.93. This simultaneous increase in average performance and decrease in variability indicates that the educational intervention, particularly the use of MEE, not only enhanced overall understanding but also led to more consistent outcomes across the entire student group.

TABLE 2 | Descriptive statistics for pre-test, inter-test, and post-test scores.

Test	Mean	Median	Mode	Standard deviation
Pre-test	5.64	6.00	5.00	1.19
Inter-test	6.05	6.00	6.00	1.29
Post-test	9.04	9.00	9.00	0.93

The evolution of central tendency measures further corroborates this trend. The median and mode, initially at 6.00 and 5.00 respectively in the pre-test, both increased to 9.00 in the post-test. This uniform improvement across different statistical measures underscores the broad-based impact of the interventions on student performance.

The trajectory of these statistics from pre-test through post-test clearly demonstrates the effectiveness of the implemented educational strategies. While traditional methods used between the pre-test and inter-test yielded only slight improvements, the innovative approaches introduced before the post-test, primarily MEE, appear to have significantly enhanced student engagement and comprehension.

This comprehensive improvement in both average scores and score consistency suggests that MEE was particularly effective in fostering a deeper and more uniform understanding of the subject matter across the student cohort. The data strongly indicates that such innovative, interactive teaching tools can lead to substantial and homogeneous improvements in student learning outcomes, especially in complex subjects like archaeological stratigraphy.

3.2 | Question Mapping Analysis

To evaluate students' comprehension of fundamental archaeological concepts, our study implemented a thorough question mapping strategy across three assessment phases: Pre-test, inter-test, and post-test. This approach encompassed key topics such as the principle of superposition, relative dating and stratigraphic application, stratigraphic units and archaeological context, the significance of stratigraphy and context, and the principle of original horizontality and stratigraphic correlation. The relationships between questions addressing these concepts across the three tests were previously outlined in Table 2. By strategically aligning questions for each concept throughout all three assessments, we enabled a precise tracking of student performance evolution. The subsequent statistical analysis of these mapped questions unveiled marked improvements in student understanding, with the most significant progress observed following the introduction of MEE.

Independent samples *t*-tests were conducted to compare student responses across different tests, with comparisons made in pairs: first between the Pre-test and the Inter-test, and then between the Inter-test and the Post-test. These statistical tests (see Table 3) enabled the evaluation of whether differences in mean scores were statistically significant, providing an objective measure of the effectiveness of educational interventions between each phase of the experiment.

This clear and detailed structure not only facilitates understanding of the study design but also underscores the systematic methodology used to evaluate the effectiveness of teaching strategies, which is essential for the validity and reproducibility of the study.

The findings indicate that traditional teaching methods yielded moderate initial improvements. However, the integration of MEE between the Inter-test and Post-test phases led to significantly more pronounced and impactful advancements. The clear statistical significance and the directionality of the *t*-statistics across various themes reflect a meaningful and progressive enhancement in student performance. Notably, these improvements became substantially more significant following the adoption of MEE, underscoring the efficacy of incorporating gamification and innovative technologies into educational strategies. This adaptation not only aligns with evolving pedagogical trends but also effectively meets learning objectives. To ensure these gains are maintained and built upon, it is recommended to conduct a further analysis of effect sizes and to continuously monitor and evaluate student progress. This ongoing assessment will help sustain the positive impacts observed and refine future teaching interventions.

3.3 | Statistical Analysis and Results: ANOVA, PCA, and Qualitative Feedback

The comprehensive analysis of student performance across the three test phases (pre-test, inter-test, and post-test) revealed

TABLE 3 | Statistical analysis of student performance on archaeological concepts.

Concept	Test compar.	*T*-stat.	*p*	Interpretation
Principle of superposition	Pre vs. Inter	−2.809	0.0056	Significant improvement
	Inter vs. Post	−3.333	0.0011	Significant improvement
Relative dating and use of stratigraphy	Pre vs. Inter	0.329	0.743	No significant change
	Inter vs. Post	−2.158	0.033	Significant improvement
Stratigraphic unit and archaeological context	Pre vs. Inter	0.168	0.867	No significant improvement
	Inter vs. Post	−5.702	6.36e-08	Substantial improvement
Importance of stratigraphy and context	Pre vs. Inter	−1.860	0.0649	Marginal improvement (not statistically significant)
	Inter vs. Post	−2.138	0.0341	Statistically significant improvement
Principle of original horizontality and stratigraphic correlation	Pre vs. Inter	0.374	0.709	No significant impact
	Inter vs. Post	−2.557	0.0116	Marked improvement

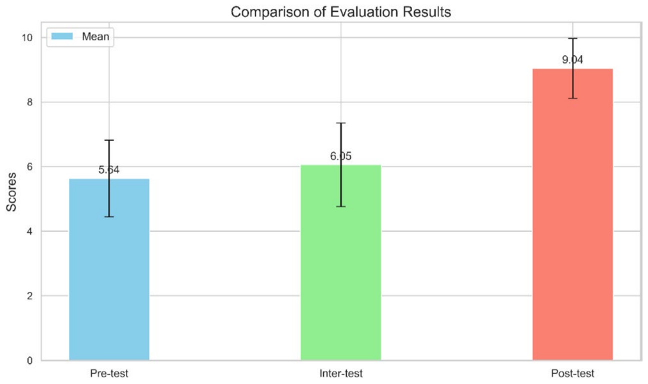

FIGURE 4 | Comparison of evaluation results by mean.

significant and meaningful changes in learning outcomes. These changes were evident in both the descriptive statistics and the inferential statistical analyses conducted.

A one-way ANOVA yielded a substantial F-statistic of 194.1811 with a p-value of 3.0981e-49, which is well below the conventional significance level of 0.05. This result strongly indicates statistically significant differences among the means of the three tests, suggesting that the educational interventions, particularly the introduction of MEE, had a measurable and meaningful impact on student performance.

The descriptive statistics (see Figure 4) provide a clear picture of the progression in student understanding. In the Pre-test, students demonstrated a baseline level of knowledge with a mean score of 5.64 ($SD = 1.19$). The Inter-test, following traditional teaching methods, showed a modest improvement with a mean score of 6.05 ($SD = 1.29$). However, the most striking change was observed in the Post-test, where the mean score increased dramatically to 9.04 ($SD = 0.93$).

3.3.1 | Pairwise Comparisons and Intervention Effectiveness

To gain more detailed insights into the differences between specific test pairs, we conducted Tukey's HSD post hoc test and paired t-tests. These analyses revealed a nuanced picture of the intervention's effectiveness at different stages.

Interestingly, there was no statistically significant difference between the pre-test and inter-test scores ($t = -1.9095$, $p = 6.0133\text{e-}02$). This suggests that the traditional teaching methods employed between these two tests, while producing a slight increase in mean scores, did not lead to a statistically significant improvement in student performance.

In contrast, highly significant differences were observed between the Inter-test and Post-test ($t = -14.9704$, $p = 5.9961\text{e-}24$), as well as between the Pre-test and Post-test ($t = -19.0936$, $p = 4.1740\text{e-}30$). These results indicate that the most substantial improvement in student performance occurred after the Inter-test, coinciding with the implementation of MEE as the main educational intervention.

3.3.2 | Magnitude of Improvement and Practical Significance

The magnitude of improvement was substantial, as evidenced by both statistical and practical measures. The mean score increased by only 0.41 points from the pre-test to the inter-test, but then dramatically by 2.99 points from the Inter-test to the Post-test, resulting in a total average improvement of 3.41 points from the initial to the final evaluation.

Moreover, the effect size, calculated using Cohen's d, was 3.1934 for the pre-test to post-test comparison. This is considered a large effect size and indicates substantial practical significance. This large effect size underscores the meaningful impact of the educational interventions, particularly MEE, on student learning outcomes in archaeological stratigraphy.

3.3.3 | PCA

The PCA of the archaeological stratigraphy test scores reveals significant insights into the learning progression of students across the pre-test, inter-test, and post-test phases. The first two principal components, accounting for approximately 34% of the total variance (PC1: 20.74%, PC2: 13.34%), provide a clear visual representation of the students' learning journey.

The scatter plot (see Figure 5) of these two components shows a distinct separation between the three test phases, with a notable progression from left to right. This pattern strongly indicates that students' understanding of archaeological stratigraphy concepts evolved substantially throughout the course. The pre-test scores, clustered predominantly on the left side of the plot, represent the initial knowledge baseline. The Inter-test scores, slightly shifted towards the right, suggest a modest improvement following traditional teaching methods. However, the most striking feature is the clear separation and rightward shift of the Post-test scores, which demonstrate the profound impact of the MEE intervention on student learning. Moreover, three ellipses have been drawn to make easier the groups recognition. These ellipses provide a visual representation of the distribution of each group, making it easier to see how the groups are separated and how their distributions change across the tests.

The condensed grouping of post-test scores not only indicates improved performance but also suggests a more uniform understanding across the student cohort. This observation aligns with the raw data, which shows a significant increase in mean scores and a decrease in standard deviation for the post-test. Such homogeneity in final scores implies that MEE was effective in bridging knowledge gaps among students, potentially offering a more inclusive learning experience that benefits learners across different initial skill levels.

Examining the loadings (see Figure 6) of individual questions on the principal components provides further insights. The first principal component, with its consistently negative loadings (except for Q1), appears to represent an overall measure of understanding across all stratigraphy concepts. The higher absolute loadings for questions 3, 4, and 5 on this component suggest that

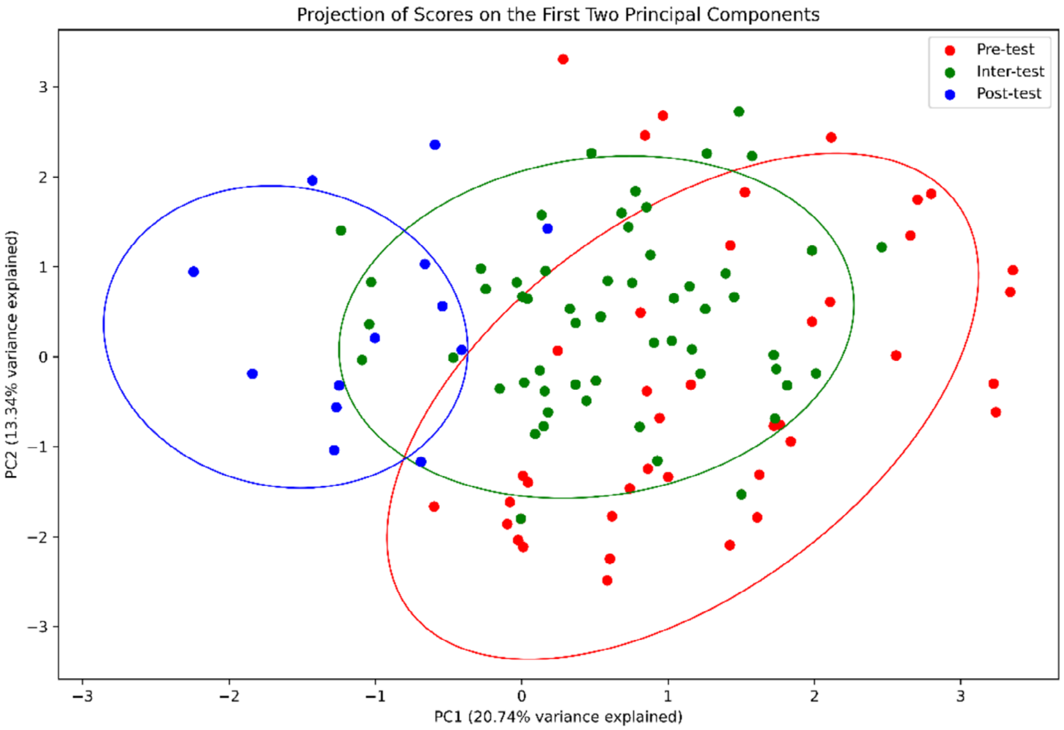

FIGURE 5 | PCA scatter plot. The confidence ellipses are drawn with a 95% confidence level (2 standard deviations).

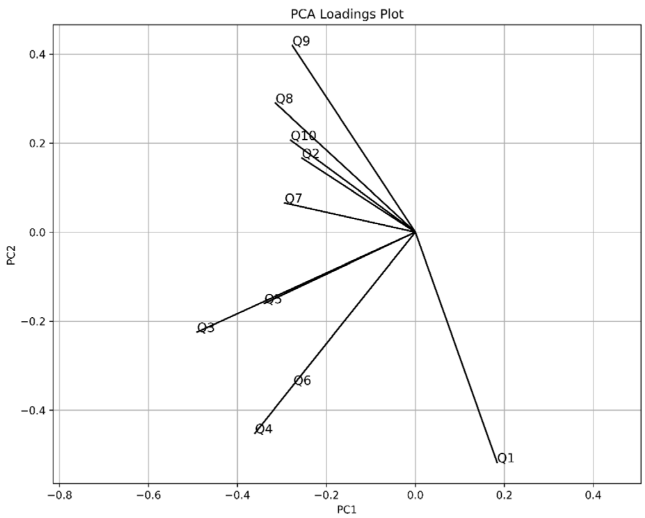

FIGURE 6 | Loadings plot.

these questions were particularly effective in discriminating between different levels of overall understanding.

The second principal component, characterised by a mix of positive and negative loadings, likely captures more nuanced aspects of learning. The strong positive loading of Q9 and negative loading of Q1 on this component might indicate a contrast between different types of knowledge or skills within the domain of archaeological stratigraphy. This component could be revealing how students' understanding of certain concepts evolved differently compared to others throughout the learning process.

The gradual decrease in variance (see Figure 7) explained by subsequent components underscores the multidimensional nature of learning archaeological stratigraphy. It suggests that while there are dominant patterns in how students learn these concepts, there are also multiple, more subtle aspects to their understanding that evolve in various ways.

The effectiveness of MEE is particularly evident when comparing the Inter-test and Post-test results. The more pronounced separation between these two phases in the PCA plot, compared to the separation between Pre-test and Inter-test, corroborates the raw score data. It demonstrates that while traditional teaching methods led to some improvement, the introduction of MEE resulted in a more substantial and consistent enhancement of student understanding.

In conclusion, the PCA results, supported by the raw test scores, provide compelling evidence for the efficacy of MEE in teaching archaeological stratigraphy. The analysis reveals not just an improvement in overall scores, but a transformation in how students understand and interrelate various concepts within the subject. The clear progression and clustering observed in the PCA plot, combined with the improvements seen in individual question responses, suggest that MEE offers a powerful tool for enhancing both the depth and consistency of student learning in this complex field. These findings have significant implications for archaeological education, pointing towards the potential of interactive, gamified learning environments to revolutionise how complex spatial and temporal concepts are taught and understood.

3.4 | Qualitative Feedback

To gain deeper insights into students' experiences, we employed a mixed-methods approach for collecting qualitative feedback.

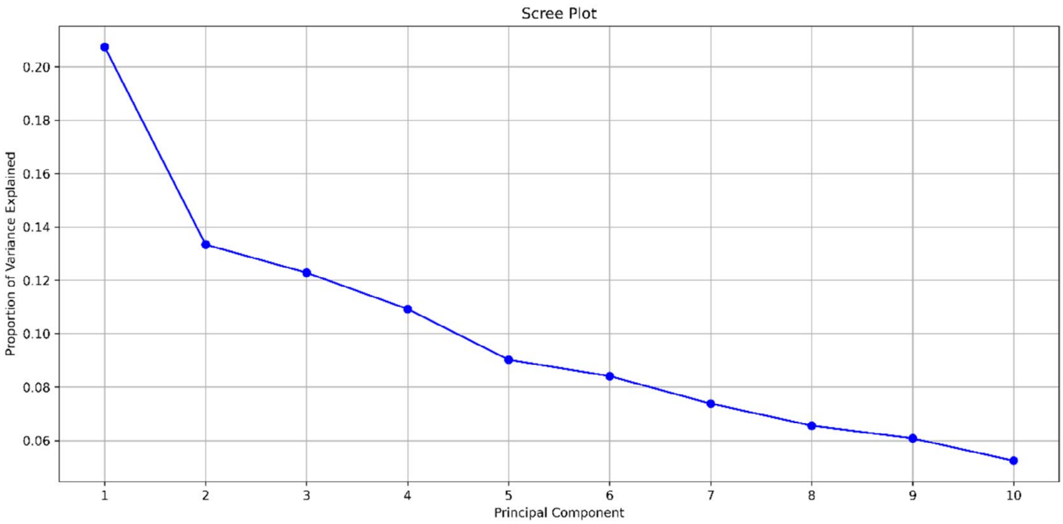

FIGURE 7 | Variance explained by principal component.

TABLE 4 | Some of the most repeated ideas on the qualitative feedback.

Theme	Positive feedback	Suggestions for improvement
Engagement	"Minecraft was fun and made learning easy."	"More guided tasks would help."
Understanding	"I could see the layers and understand better."	"Sometimes it was hard to know what to do next."
Practical Application	"Felt like real excavation, very helpful."	"A bit more time to practise would be great."
Overall Experience	"Preferred Minecraft over traditional lectures."	"Combining both methods might be the best approach."

Following the completion of the Post-test, students were asked to complete an anonymous online survey comprising open-ended questions about their learning experiences with both traditional methods and MEE. Additionally, we conducted semi-structured interviews with a subset of 15 randomly selected students to explore their perceptions in greater depth (Braun and Clarke 2006; Clarke et al. 2015). The responses from both the surveys and interviews were systematically coded using thematic analysis to identify recurring themes and patterns in students' feedback.

The themes presented in Table 4 emerged from this rigorous coding process, with 'Engagement' and 'Understanding' being the most frequently mentioned aspects across all responses. Notably, these qualitative findings align closely with our quantitative results, particularly the significant improvement in test scores and reduced variability observed in the Post-test. For instance, the theme of 'Understanding' corroborates the quantitative data showing improved performance, with students frequently mentioning how MEE's visual and interactive nature enhanced their grasp of stratigraphic concepts. One student remarked, "Being able to actually see and manipulate the layers in Minecraft made it so much easier to understand how stratigraphy works in real life."

The 'Engagement' theme provides valuable context for the increased consistency in Post-test scores. Many students reported feeling more motivated and attentive when using MEE, which may explain why the intervention benefited students across different ability levels. As one participant noted, "I usually struggle to stay focused during lectures, but with Minecraft, I was completely immersed in the learning process." This increased engagement across the board could account for the reduced standard deviation observed in the post-test scores.

These qualitative insights not only support our quantitative findings but also offer explanations for the observed improvements, highlighting the multifaceted benefits of incorporating interactive, technology-based learning tools like MEE in archaeology education.

4 | Discussion and Conclusion

This study set out to evaluate the effectiveness of MEE as an innovative tool for teaching archaeological stratigraphy compared to traditional teaching methods. The results, including our PCA, provide compelling evidence for the effectiveness of MEE in enhancing student understanding and engagement with complex archaeological concepts.

The quantitative analysis, now augmented by PCA, revealed a striking improvement in student performance following the implementation of MEE. The one-way ANOVA results (F-statistic:

194.1811, p-value: 3.0981e-49) unequivocally demonstrate that the differences in mean scores across the three tests were not due to chance. This finding strongly supports the hypothesis that MEE significantly impacted student learning outcomes.

The progression of mean scores from pre-test (5.64) to inter-test (6.05) and finally to post-test (9.04) illustrates the differential impact of traditional teaching methods versus MEE. The PCA results further corroborate this progression, showing a clear separation between the three test phases in the scatter plot of the first two principal components. The modest improvement between Pre-test and inter-test, visible in both mean scores and the PCA plot, suggests that while traditional methods do contribute to learning, their impact is limited. In contrast, the dramatic increase in scores and the distinct rightward shift of Post-test data points in the PCA plot following the introduction of MEE indicates its superior effectiveness in facilitating student understanding of archaeological stratigraphy.

Particularly noteworthy is the reduction in standard deviation from pre-test (1.19) to post-test (0.93), coupled with the increase in mean score. This simultaneous improvement in both average performance and consistency is visually represented in the PCA plot, where Post-test scores cluster more tightly. This suggests that MEE is effective across a broad spectrum of students, potentially narrowing the gap between high and low performers. This finding has significant implications for inclusive education practises in archaeology.

The large effect size (Cohen's $d = 3.1934$) between Pre-test and Post-test scores further emphasises the practical significance of these improvements. In educational research, such a large effect size is rare and underscores the transformative potential of MEE in archaeology education.

The question mapping analysis, complemented by the PCA loadings, provided granular insights into how MEE impacted understanding of specific archaeological concepts. The consistent pattern of significant improvements from Inter-test to Post-test across all concepts (Principle of Superposition, Relative Dating, Stratigraphic Units, etc.) suggests that MEE's benefits are not limited to a narrow aspect of stratigraphy but encompass a broad range of fundamental archaeological principles.

The PCA loadings reveal that questions 3, 4, and 5 had the highest absolute loadings on the first principal component, indicating their importance in discriminating overall understanding. This aligns with the particularly striking improvement observed in understanding of stratigraphic units and archaeological context (t-statistic: −5.702, p-value: 6.36e-08). The PCA results suggest that MEE's visual and interactive nature is especially effective in conveying spatial and contextual relationships within archaeological sites, a crucial skill for aspiring archaeologists.

The lack of significant improvement in most concepts between Pre-test and Inter-test, further highlighted by the minimal separation in the PCA plot, underscores the limitations of traditional teaching methods, especially for complex, spatially-oriented concepts in archaeology. This finding challenges the status quo in archaeological education and calls for a reconsideration of pedagogical approaches in the field.

The qualitative feedback provided crucial context for understanding the mechanisms behind MEE's efficiency. The recurring themes of enhanced engagement and improved understanding align closely with the quantitative results, offering explanatory power to the observed improvements.

Students' reports of increased motivation and immersion when using MEE suggest that the platform's gamification elements play a crucial role in its effectiveness. By making the learning process more enjoyable and interactive, MEE appears to overcome common barriers to engagement with complex archaeological concepts. This increased engagement likely contributes to the improved retention and understanding reflected in the Post-test scores.

The theme of improved understanding, particularly through visual and interactive experiences, supports theories of experiential learning in archaeology. MEE's ability to allow students to "see and manipulate" stratigraphic layers provides a virtual hands-on experience that traditional methods struggle to replicate. This finding suggests that MEE may be particularly valuable in contexts where access to physical archaeological sites is limited or impractical.

4.1 | Pedagogical Implications

The results of this study have significant implications for archaeological education. The clear superiority of MEE over traditional methods in improving student understanding challenges long-standing pedagogical approaches in the field. It suggests that incorporating interactive, technology-based tools like MEE should be a priority in curriculum development for archaeology programmes.

However, the qualitative feedback also indicates that a complete replacement of traditional methods may not be optimal. Some students suggested that "combining both methods might be the best approach." This insight points towards a blended learning model, where MEE is used to complement and enhance traditional teaching methods rather than replace them entirely.

The improved consistency in student performance after using MEE also has implications for addressing educational inequities in archaeology. By providing a standardised, interactive learning environment, MEE may help level the playing field for students who might struggle with traditional lecture-based learning.

The implementation of MEE in archaeological education offers several significant advantages. Firstly, it enhances visualisation by providing dynamic, three-dimensional representations of archaeological concepts, which are particularly beneficial for understanding complex spatial relationships in stratigraphy. The gamification elements inherent in MEE foster increased engagement, motivating students and leading to improved learning outcomes. Moreover, MEE offers a unique opportunity for experiential learning, allowing students to interact with archaeological concepts through virtual hands-on experiences that are difficult to replicate using traditional teaching methods. The platform also standardises the learning experience, potentially reducing disparities in educational quality and access across different institutions. Finally, the scalability of MEE modules, once developed, allows for easy distribution and implementation across various educational settings, potentially democratising access to high-quality archaeological education on a global scale. These combined advantages position MEE as a powerful tool for revolutionising archaeological education and training.

Despite the promising results of this study, it is important to acknowledge several limitations that warrant consideration and further investigation. The sample size and diversity of the study participants represent a significant constraint. Specifically, this research was conducted with a cohort of 74 undergraduate students enrolled in an archaeological stratigraphy course at Universitat de València during the 2nd semester 2024. This sample, while providing valuable insights, is limited in its size and demographic representation. The participants were primarily 20–25 years old, with 63% identifying as male, and 95% coming from the Humanities branch in secondary school. This homogeneity in the sample may limit the generalisability of our findings to broader and more diverse student populations or to students in different academic contexts. Future studies should aim to include a larger and more diverse sample, encompassing students from various universities, academic levels, and cultural backgrounds to enhance the robustness and applicability of the findings. To establish the generalisability of these findings, it is crucial to replicate the study with larger and more diverse student populations across different educational and cultural contexts. Additionally, the focus on immediate learning outcomes leaves questions about long-term knowledge retention unanswered. Future longitudinal studies are necessary to assess whether the knowledge gained through MEE is retained over extended periods and how it compares to traditional teaching methods in this aspect.

Technical barriers pose another limitation, as the successful implementation of MEE requires access to appropriate technology and technical support, which may not be universally available, potentially exacerbating educational inequalities if not addressed. There is also a risk of potential overreliance on technology, where an overemphasis on digital tools like MEE could inadvertently lead to the neglect of important traditional archaeological skills that remain crucial in the field. Lastly, while MEE proved highly effective for teaching stratigraphy, its efficacy in other areas of archaeology education remains untested. Further research is needed to explore its applicability and effectiveness across the broader spectrum of archaeological studies, including artefact analysis, site surveying, and theoretical concepts. Addressing these limitations in future research will be crucial for fully understanding the potential and appropriate integration of MEE in archaeological education.

4.2 | Future Perspectives

The success of MEE in teaching stratigraphy opens up exciting possibilities for its application in other areas of archaeology:

- Site reconstruction: MEE could be used to create virtual reconstructions of archaeological sites, allowing students to explore and interact with past environments.
- Artefact analysis: Interactive modules could be developed for teaching artefact classification and analysis techniques.
- Excavation simulation: MEE could provide a safe, cost-effective way to teach excavation methods and decision-making skills.
- Cultural heritage preservation: Virtual models in MEE could be used to teach and promote cultural heritage preservation techniques.
- Simulation of real archaeological sites representing different excavation conditions (i.e., open fields, caves, underwater intervention, surveys).

Standardisation of MEE modules for archaeology education presents both opportunities and challenges. On one hand, standardised modules could ensure consistency in the quality of education across institutions. On the other, care must be taken to allow for regional variations in archaeological practise and to avoid a one-size-fits-all approach.

Future research should focus on developing best practises for integrating MEE into archaeology curricula, assessing its effectiveness across different cultural and educational contexts, and exploring its potential in professional archaeological training and public outreach.

Moreover, from a psychometric perspective, future research should focus on developing and validating standardised assessment tools for archaeological stratigraphy education. While our current study provided promising initial results, a comprehensive psychometric validation process would be valuable for developing standardised instruments. This could include:

- Large-scale validation studies across multiple institutions to establish reliability and validity metrics.
- Development of item banks with known psychometric properties for assessing stratigraphic understanding
- Longitudinal studies to evaluate test–retest reliability and knowledge retention
- Cross-cultural validation to ensure assessment tools are effective across different educational contexts

Such psychometrically validated tools would not only strengthen future research in this field but also provide educators with reliable instruments for assessing student learning in archaeological stratigraphy education.

In conclusion, this study provides strong evidence for the potential and effectiveness of Minecraft: Education Edition as a tool for teaching archaeological stratigraphy. Its ability to significantly improve student understanding and engagement suggests that it has the potential to revolutionise archaeology education. While challenges remain in its implementation and standardisation, the benefits of MEE in archaeology education are clear and compelling. As we move forward, the integration of such innovative, technology-based tools in archaeology education will likely play a crucial role in preparing the next generation of archaeologists for the challenges and opportunities of 21st-century archaeology. In conclusion, the adoption of MEE in teaching archaeological stratigraphy represents a significant advancement in educational methodologies. It allows for a more holistic and engaging learning experience that not only improves outcomes but also ignites a passion for learning among students. As educational landscapes continue to evolve, the integration of such innovative tools will be crucial in preparing students not just to learn, but to thrive in a complex and rapidly changing world. The findings from this study provide a strong foundation for further exploration into the use of digital tools in education and highlight the need for ongoing research to optimise these technologies for educational use.

In moving forward, it will be important for educational institutions to consider the integration of similar digital tools across various disciplines. The success of MEE in teaching archaeological stratigraphy opens the door for its application in other areas of study, potentially revolutionising the way educational content is delivered and experienced across the board. As we continue to explore the boundaries of what is possible in education through technology, the lessons learned from this study will serve as valuable guideposts for future innovations.

In conclusion, this study, bolstered by the insights from PCA, provides strong evidence for the effectiveness of Minecraft: Education Edition as a tool for teaching archaeological stratigraphy. Its ability to significantly improve student understanding and engagement, as demonstrated by both traditional statistical measures and advanced multivariate analysis, suggests that it has the potential to revolutionise archaeology education. While challenges remain in its implementation and standardisation, the benefits of MEE in archaeology education are clear and compelling. As we move forward, the integration of such innovative, technology-based tools in archaeology education, coupled with sophisticated analytical approaches like PCA, will likely play a crucial role in preparing the next generation of archaeologists for the challenges and opportunities of 21st-century archaeology.

Author Contributions

Joaquín Jiménez-Puerto: conceptualization, methodology, software, data curation, investigation, visualization, resources, writing – original draft, writing – review and editing. **Gianni Gallello:** conceptualization, formal analysis, validation, writing – review and editing.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that supports the findings of this study are available in the Supporting Information of this article.

References

Abdi, H., and L. J. Williams. 2010. “Tukey's Honestly Significant Difference (HSD) Test.” *Encyclopedia of Research Design* 3, no. 1: 1–5.

AlJanah, S., P. S. Teh, J. Y. Tay, O. Aiyenitaju, and R. Nawaz. 2023. “Minecraft as a Tool to Enhance Engagement in Higher Education.” In *Research and Innovation Forum 2022*, edited by A. En Visvizi, O. Troisi, and M. Grimaldi, 465–476. Springer International Publishing. https://doi.org/10.1007/978-3-031-19560-0_38.

All, A., E. P. N. Castellar, and J. Van Looy. 2016. “Assessing the Effectiveness of Digital Game-Based Learning: Best Practices.” *Computers & Education* 92: 90–103.

Andersen, R., and M. Rustad. 2022. “Using Minecraft as an Educational Tool for Supporting Collaboration as a 21st Century Skill.” *Computers and Education Open* 3: 100094. https://doi.org/10.1016/j.caeo.2022.100094.

Anderson, L. W., and D. R. Krathwohl. 2001. *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives: Complete Edition*. Addison Wesley Longman, Inc.

Bandalos, D. 2018. *Measurement Theory and Applications for the Social Sciences*. Guilford Publications.

Boom, K. H. J., C. E. Ariese, B. Van Den Hout, A. A. A. Mol, and A. Politopoulos. 2020. “Teaching Through Play: Using Video Games as a Platform to Teach About the Past.” In *Communicating the Past in the Digital Age: Proceedings of the International Conference on Digital Methods in Teaching and Learning in Archaeology*, edited by S. En Hageneuer, 27–44. Ubiquity Press.

Braun, V., and V. Clarke. 2006. “Using Thematic Analysis in Psychology.” *Qualitative Research in Psychology* 3, no. 2: 77–101.

Callaghan, N. 2016. “Investigating the Role of Minecraft in Educational Learning Environments.” *Educational Media International* 53, no. 4: 244–260. https://doi.org/10.1080/09523987.2016.1254877.

Carver, M. 2013. *Archaeological Investigation*. Routledge.

Champion, E. 2016. *Critical Gaming: Interactive History and Virtual Heritage*. Routledge. https://doi.org/10.4324/9781315574981.

Charsky, D. 2010. “From Edutainment to Serious Games: A Change in the Use of Game Characteristics.” *Games and Culture* 5, no. 2: 177–198.

Chenrai, P. 2021. “Case Study on Geoscience Teaching Innovation: Using 3D Printing to Develop Structural Interpretation Skill in Higher Education Levels.” *Frontiers in Earth Science* 8: 590062. https://doi.org/10.3389/feart.2020.590062.

Clarke, V., V. Braun, and N. Hayfield. 2015. “Thematic Analysis.” In *Qualitative Psychology: A Practical Guide to Research Methods*, edited by J. A. Smith, 222–248. 3rd ed. SAGE Publications.

Cohen, L., L. Manion, and K. Morrison. 2002. *Research Methods in Education*. Routledge.

Creswell, J. W., and J. D. Creswell. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage Publications.

Crocker, L., and J. Algina. 1986. *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston.

Deterding, S., D. Dixon, R. Khaled, and L. Nacke. 2011. “From Game Design Elements to Gamefulness: Defining Gamification.” In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, 9–15. ACM. https://doi.org/10.1145/2181037.2181040.

Ellenberger, K. 2017. “Virtual and Augmented Reality in Public Archaeology Teaching.” *Advances in Archaeological Practice* 5, no. 3: 305–309. https://doi.org/10.1017/aap.2017.20.

Field, A. 2024. *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Limited.

Fuchs, M., S. Fizek, and P. Ruffino, eds. 2014. *Rethinking Gamification*. Meson Press. https://doi.org/10.14619/001.

Gallello, G., and S. Machause. 2024. “Diseño, desarrollo y aplicación de ejercicios LEGO para la comprensión de una secuencia estratigráfica en arqueología.” X Congreso de Innovación Educativa y Docencia en Red, 11-12 Junio 2024.

Garstki, K. J., C. Larkee, and J. LaDisa. 2019. “A Role for Immersive Visualization Experiences in Teaching Archaeology.” *Studies in Digital Heritage* 3, no. 1: 46–59. https://doi.org/10.14434/sdh.v3i1.25145.

Gee, J. P. 2003. “What Video Games Have to Teach Us About Learning and Literacy.” *Computers in Entertainment* 1, no. 1: 20. https://doi.org/10.1145/950566.950595.

Gelman, A. 2005. “Analysis of Variance—Why it Is More Important Than Ever.”

Hake, R. R. 1998. “Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses.” *American Journal of Physics* 66, no. 1: 64–74.

Hamari, J., J. Koivisto, and H. Sarsa. 2014. “Does Gamification Work?–A Literature Review of Empirical Studies on Gamification.” 47th Hawaii International Conference on System Sciences. 3025-3034.

Hambleton, R. K., and R. W. Jones. 1993. “Comparison of Classical Test Theory and Item Response Theory and Their Applications to Test Development.” *Educational Measurement: Issues and Practice* 12, no. 3: 38–47.

Harris, E. C. 1979a. *Principles of Archaeological Stratigraphy*. Elsevier.

Harris, E. C. 1979b. “The Laws of Archaeological Stratigraphy.” *World Archaeology* 11, no. 1: 111–117.

Herrera, J. S., and E. M. Riggs. 2013. “Identifying Students' Conceptions of Basic Principles in Sequence Stratigraphy.” *Journal of Geoscience Education* 61, no. 1: 89–102.

Jolliffe, I. T. 2002. *Principal Component Analysis for Special Types of Data*. Springer.

Kokotsaki, D., V. Menzies, and A. Wiggins. 2016. “Project-Based Learning: A Review of the Literature.” *Improving Schools* 19, no. 3: 267–277.

Kolb, D. A. 1984. *Experiential Learning: Experience as the Source of Learning and Development*. Prentice-Hall.

Kolb, D. A. 2014. *Experiential Learning: Experience as the Source of Learning and Development*. FT Press.

Kreager, B., N. LaDue, and T. Shipley. 2023. “Conceptual Understanding of Sequence Stratigraphy.” *Journal of Geoscience Education* 71, no. 4: 507–524.

Laurillard, D. 2013. *Rethinking University Teaching: A Conversational Framework for the Effective Use of Learning Technologies*. Routledge.

Lave, J., and E. Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.

Lee, S., M. C. Harrison, G. Pell, and C. L. Robinson. 2008. “Predicting Performance of First Year Engineering Students and the Importance of Assessment Tools Therein.” *Engineering Education* 3, no. 1: 44–51.

Liu, Q., W. Peng, F. Zhang, R. Hu, Y. Li, and W. Yan. 2016. “The Effectiveness of Blended Learning in Health Professions: Systematic Review and Meta-Analysis.” *Journal of Medical Internet Research* 18, no. 1: e2.

Mayer, R. E. 2024. “The Past, Present, and Future of the Cognitive Theory of Multimedia Learning.” *Educational Psychology Review* 36, no. 1: 8.

Microsoft. 2024. “Minecraft: Education Edition Usage Statistics.”

Mojang Studios. 2024. *Minecraft Education Edition*. Microsoft Education. https://education.minecraft.net/.

Morgan, C. 2019. “Avatars, Monsters, and Machines: A Cyborg Archaeology.” *European Journal of Archaeology* 22, no. 3: 324–337. https://doi.org/10.1017/eaa.2019.22.

Nixon, L., and S. Price. 2004. “Paper, Video, Internet: New Technologies for Research and Teaching in Archaeology: The Sphakia Survey.” *Journal of Interactive Media in Education* 2004, no. 2: 26. https://doi.org/10.5334/2004-16.

Nkadimeng, M., and P. Ankiewicz. 2022. “The Affordances of Minecraft Education as a Game-Based Learning Tool for Atomic Structure in Junior High School Science Education.” *Journal of Science Education and Technology* 31, no. 5: 605–620. https://doi.org/10.1007/s10956-022-09981-0.

OECD. 2019. “OECD Future of Education and Skills 2030: OECD Learning Compass 2030.” A Series of Concept Notes.

Papert, S. A. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books.

Piaget, J. 1977. *The Development of Thought: Equilibration of Cognitive Structures*. Viking.

Politopoulos, A., A. A. A. Mol, K. H. J. Boom, and C. E. Ariese. 2019. “History Is Our Playground: Action and Authenticity in Assassin's Creed: Odyssey.” *Advances in Archaeological Practice* 7, no. 3: 317–323. https://doi.org/10.1017/aap.2019.30.

Prensky, M. 2001. “Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently.” *On the Horizon* 9, no. 6: 1–6. https://doi.org/10.1108/10748120110424843.

Price, L. R. 2017. *Psychometric Methods: Theory Into Practice*. Guilford Press.

Renfrew, C., and P. G. Bahn. 2007. *Archaeology Essentials: Theories, Methods, and Practice*. Thames & Hudson.

Schrier, K. 2018. “Guiding Questions for Game-Based Learning.” In *Second Handbook of Information Technology in Primary and Secondary Education*, edited by J. Voogt, G. Knezek, R. Christensen, and K. W. Lai, 1–20. Springer International Publishing. https://doi.org/10.1007/978-3-319-53803-7_59-1.

Sireci, S., and J.-L. Padilla. 2014. “Validating Assessments: Introduction to the Special Section.” *Psicothema* 1: 97–99.

Slattery, E. J., D. Butler, M. O'Leary, and K. Marshall. 2023. “Teachers' Experiences of Using Minecraft Education in Primary School: An Irish Perspective.” *Irish Educational Studies* 43, no. 4: 1–20. https://doi.org/10.1080/03323315.2023.2185276.

Sullivan, G. M., and R. Feinn. 2012. “Using Effect Size—Or Why the P Value Is Not Enough.” *Journal of Graduate Medical Education* 4, no. 3: 279–282.

Waagen, J. 2019. “New Technology and Archaeological Practice. Improving the Primary Archaeological Recording Process in Excavation by Means of UAS Photogrammetry.” *Journal of Archaeological Science* 101: 11–20. https://doi.org/10.1016/j.jas.2018.10.011.

Zimmerman, D. W. 1997. “Teacher's Corner: A Note on Interpretation of the Paired-Samples t Test.” *Journal of Educational and Behavioral Statistics* 22, no. 3: 349–360.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.