Investigating the Benefits of Physical Models for Anatomical Education in Augmented Reality

Matthew Wragg
Department of Computer Science
University of Bath
Bath, United Kingdom
mw2399@bath.ac.uk

Raj Sengupta
Royal National Hospital for Rheumatic Diseases
Bath, United Kingdom
rajsengupta@nhs.net

Dario Cazzola
Department for Health
University of Bath
Bath, United Kingdom
dc547@bath.ac.uk

Jason Alexander
Computer Science
University of Bath
Bath, United Kingdom
jma73@bath.ac.uk

Figure 1: We created two AR systems and one screen-based system as educational tools to support collaborative learning around spinal anatomy and arthritis disease progression. The tangible AR system (left) utilises an underlying physical vertebrae model held in one user's hand, and the virtual AR system (middle) displays a purely holographic vertebrae model. The screen-based system (right) uses an interactive desktop application.

Abstract

Historically, anatomical education has utilised physical models; researchers are now looking to Augmented Reality (AR) to deliver more engaging learning experiences. While there are clear educational advantages to AR, most systems lack the cognitive benefits afforded by physical models. Our work explores the potential of combining physical anatomical models and AR. We first present a design space exploring the interplay between the two. From this, we created a tangible AR system utilising a physical vertebrae model for learning spinal anatomy and axial spondyloarthritis progression. We conducted a study (n=39) to evaluate its benefits for knowledge improvement and retention, compared with a virtual AR and screen-based version. We found no difference in learning outcomes, however, the physical model improved participants' learning experience. We then conducted an expert evaluation with clinicians to explore opportunities for using tangible AR in clinical practice. Results highlight potential benefits for patient understanding, and challenges surrounding accessibility.

CCS Concepts

• **Human-centered computing → Mixed / augmented reality**; • **Applied computing → Health informatics**; **Interactive learning environments**.

Keywords

augmented reality, tangible interaction, physical models, anatomy, education

ACM Reference Format:
Matthew Wragg, Raj Sengupta, Dario Cazzola, and Jason Alexander. 2025. Investigating the Benefits of Physical Models for Anatomical Education in Augmented Reality. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3706598.3713733

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713733

1 Introduction

For hundreds of years, the study of anatomy has utilised physical learning modalities, such as cadaveric dissection and physical anatomical models, to provide learners with tangible and realistic 3D representations of anatomical structures and processes. Research has shown that interacting with physical objects can provide significant cognitive benefits, and that within anatomical education, physical models can offer significant educational value [20, 35, 48, 52].

As immersive technologies such as Augmented Reality (AR) have matured, it is now possible to explore virtual equivalents of these physical models in interactive and engaging environments. Such systems retain the inherent spatial realism of physical representations, while facilitating complex and *adaptable* visualisations that would not be possible within the material and mechanical constraints of physical objects. Despite these benefits, virtual models are not able to replicate the tangible affordances of their physical equivalents, losing any benefits of physical interaction.

The incorporation of physical interaction and manipulation techniques within AR, known as *tangible* AR [4], has been studied extensively [15, 28, 41]. Continuing advancements in object tracking and fabrication techniques are providing opportunities to incorporate underlying physical representations of virtual objects within AR systems. However, prior research into the use of AR for anatomical education has primarily focused on purely virtual systems with little physical interaction, and the net educational benefits of utilising physical models within AR for learning anatomy are still unknown. It is therefore unclear whether tangible AR can provide significant educational benefits over purely virtual AR for learning in the anatomy domain.

Additionally, the use of AR systems for anatomical education *in clinical practice* with patients is relatively underexplored in the literature [46], and no prior work has investigated the use of *tangible* AR specifically. Gaps in knowledge remain regarding how physical models can be utilised within AR to support anatomical education, how the addition of physical models within AR provides educational benefits compared with non-tangible AR equivalents, and the challenges associated with integrating such systems into clinical workflows. Based on these gaps in knowledge, our work is guided by the following research questions:

- ***RQ1:*** How do the affordances of anatomical models and AR, along with the interplay between them, inform the design of tangible AR systems for anatomical education?
- ***RQ2:*** How does the use of tangible AR affect immediate and long-term knowledge gains compared to non-tangible AR and screen-based systems for anatomical education?
- ***RQ3:*** How does the use of tangible AR affect learning experience compared to non-tangible AR and screen-based systems for anatomical education?
- ***RQ4:*** What opportunities and challenges exist for the use of tangible AR for anatomical education in clinical practice?

To investigate these research questions, we first present a design space based on prior work on tangible AR and anatomical education, which captures the affordances of AR and physical anatomical models, and the interplay between them. Our design space aims to classify existing systems and provide inspiration for the design of future systems through its generative power. From this design space, we created two educational AR systems. Each system presented learning material consisting of digital models and associated textual descriptions to help users understand spinal anatomy (specifically two lumbar vertebrae) and the disease progression of axial spondyloarthritis (axSpA), a form of inflammatory arthritis that primarily affects the spine and pelvic joints. One system incorporated a physical vertebral model with the learning material superimposed onto it (the **tangible AR** system). The other system presented the learning material purely as holograms with no physical component (the **virtual AR** system). In addition, we created an equivalent **screen-based** version of the system running on a desktop PC that aimed to replicate as much of the functionality of the AR systems as possible. Using these systems, we conducted two independent user studies:

The first was a comparative user study to evaluate the educational benefits of the tangible AR system. Through this study, we evaluated how the addition of a physical model within the AR system affected knowledge gain, knowledge retention, and learning experience, compared with an equivalent non-tangible AR system, and a screen-based system. 39 participants were assigned to one of three groups associated with the three systems (**tangible AR**, **virtual AR** and, **screen**). Participants completed a guided learning session before completing a series of 16 open-ended questions to assess their understanding of the material. Participants were asked to return after one week to complete the same assessment again in order to assess their long-term retention of the material. We did not find any significant benefits to learning between the different systems, however qualitative feedback from participants highlighted benefits to the overall learning experience.

In the second study, three clinicians working in axSpA care in the Royal United Hospital, Bath were invited to take part in individual guided explorations of both AR systems. Following each session, semi-structured interviews were conducted to understand their opinions of each system, and to explore potential opportunities and challenges for using such systems for patient education in clinical practice. Using thematic analysis of the interviews, we identified several notable themes including the educational benefits of the physical model, considerations regarding system usability and accessibility, and barriers for integration into clinical workflows.

In summary, our work contributes:

- ***C1:*** A design space representing the interplay between augmented reality and physical anatomical models for anatomical education, along with generative examples.
- ***C2:*** Results of a comparative study evaluating the effectiveness of tangible AR for anatomical education compared with non-tangible AR and screen-based systems.
- ***C3:*** Opportunities and challenges for the deployment of both tangible and non-tangible AR systems for anatomical education in clinical practice, identified through an expert qualitative evaluation of our system with clinicians.

2 Related Work

2.1 Augmented Reality in Anatomical Education

Particularly over the last decade, researchers have begun to explore the potential roles of immersive technologies for the delivery of anatomical education. A systematic review and meta-analysis by García-Robles et al. [14] explored the use of XR in anatomy education, looking at studies that compared XR to traditional educational approaches, such as textbooks, physical models, and cadaveric dissection. Of the 27 studies in their review, 15 focused specifically on AR technologies. For knowledge improvement, and for usefulness or perceived effectiveness, AR technologies were found to be superior to traditional approaches ($p = 0.042$), particularly when used to complement existing traditional resources. It is important to note that this review and meta-analysis analysed literature pertaining exclusively to medical students, and did not include any studies with patients or lay users. A systematic review by Urlings et al. [46] found this lack of patient-focused interventions to be representative of the wider research space. They identified few studies (n=10) that explored patient education in AR, with the available studies containing heterogeneous applications and populations which limit the generalisability of their findings.

In contrast to the findings of García-Robles et al. [14], a systematic review and meta-analysis by Bölek et al. [7] found that for AR interventions in particular, there are relatively few studies assessing learning outcomes in anatomical education. They conclude that current research does not provide sufficient evidence to suggest that AR interventions can significantly impact learning outcomes, and suggests that outcomes are impacted by learners' spatial abilities. Similar to the findings presented by Urlings et al. [46], this problem is compounded by the fact existing research contains heterogeneous AR systems, including mobile, head-mounted, and mirror-based systems, and it is still not clear which AR modality provides the greatest benefits [7]. They also highlight that while most studies in their review included motivation as an outcome measure, there is no validated method of measuring motivation for learning anatomy.

The contrasting findings reported across multiple systematic reviews and meta analyses, and the relative lack of research in this area, highlight the need for further research to evaluate the benefits of AR on learning outcomes, and overall learning experience including motivation and perceived usefulness [7].

2.2 Physical Models in Anatomical Education

Touch is a integral part of how we interact with and understand our surroundings. A great deal of previous research has shown the cognitive benefits of interactions with physical objects [2, 20, 32, 33, 42, 44]. For understanding anatomy in particular, physical models have been used as a learning tool for hundreds of years [47]. Their tangibility and close spatial correspondence to real anatomical structures has been shown to benefit spatial understanding and long-term knowledge retention [35, 52].

Estevez et al. [12] evaluated the use of 3D physical brain models to improve learning outcomes and learning experiences in neuroanatomy education. They found that the physical models were effective in improving the understanding of complex spatial relationships compared to traditional 2D cross-sectional images, and that participants perceived the experience as helpful and relatively enjoyable. Similarly, Preece et al. [35] demonstrated the benefits of a physical equine foot model for learning MRI foot anatomy compared to textbooks and computer-based 3D models. Research by Reinschluessel et al. [37] and Muender et al. [30] has evaluated the use of tangible organ-shaped controllers for manipulating virtual liver models based on medical imaging data. Studies with surgeons revealed preferences for softer, more realistic models for surgical planning in VR; high similarity between virtual and physical objects; and physical manageability over accuracy in size compared with a real liver.

Despite their usefulness, physical models are inherently limited by their relative lack of visual and spatial manipulability. While this does not limit the representation of static anatomical structures, challenges remain for representing *dynamic* processes, including complex structural changes such as bone formation and erosion.

2.3 The Intersection of Tangibility and AR

While both physical models and AR can provide benefits for anatomical education, each modality has its own distinct limitations. Fortunately, some of these limitations can be mitigated through the combination of the two modalities. Physical models provide the cognitive benefits and natural interactions of tangibility to an otherwise intangible AR system, and in turn, AR provides a means of enhancing our visuospatial perception of otherwise static models.

Prior research into the use of AR for anatomical education has typically not incorporated any sort of physical interaction. However, Cercenelli et al. [8] developed *AEducaAR*, an educational AR tool developed using Vuforia [36] for the HoloLens 2 that utilised a 3D-printed human skull model as the basis for anatomical visualisations. They conducted a study with medical students in which they evaluated the effects of their tool on learning outcomes and user experience, compared with a control group using the same 3D-printed skull and a human anatomy atlas. While they did not find any statistically significant difference in assessment scores between the two groups, the students' perception of the AR tool was positive and they found it to be more engaging and useful compared with textbook learning.

The intersection of physical models and AR still poses significant technical challenges. A key example is the challenges associated with the real-time tracking of physical models, partly due to environmental variations such as illumination and occlusion [49], which may contribute to the lack of research around its use for anatomical education. This is particularly true for off-the-shelf models that do not incorporate markers or sensors within their design that may aid with tracking and alignment. Similar challenges were highlighted by Cercenelli et al. [8] in their evaluation of their *AEducaAR* tool. They found that registration and tracking of their physical skull model was sensitive to ambient light conditions in their study environment. Studies by Barmaki et al. [3] and Bork et al. [6] have utilised an AR "magic mirror" to superimpose anatomical and radiographic visualisations directly onto the user's body. Both studies reported significant benefits to learning using the AR system compared to traditional methods. To date, we do not believe any research has explored the use of tangible AR for anatomical education compared with non-tangible AR, or how such systems affect long-term knowledge retention.

3 Design Space

To explore the combination of physical anatomical models and AR, we present a design space encompassing important design dimensions for the interplay between these two (**RQ1**). Our design space focuses on attributes and affordances of both physical anatomical models and AR that have implications for designing tangible AR

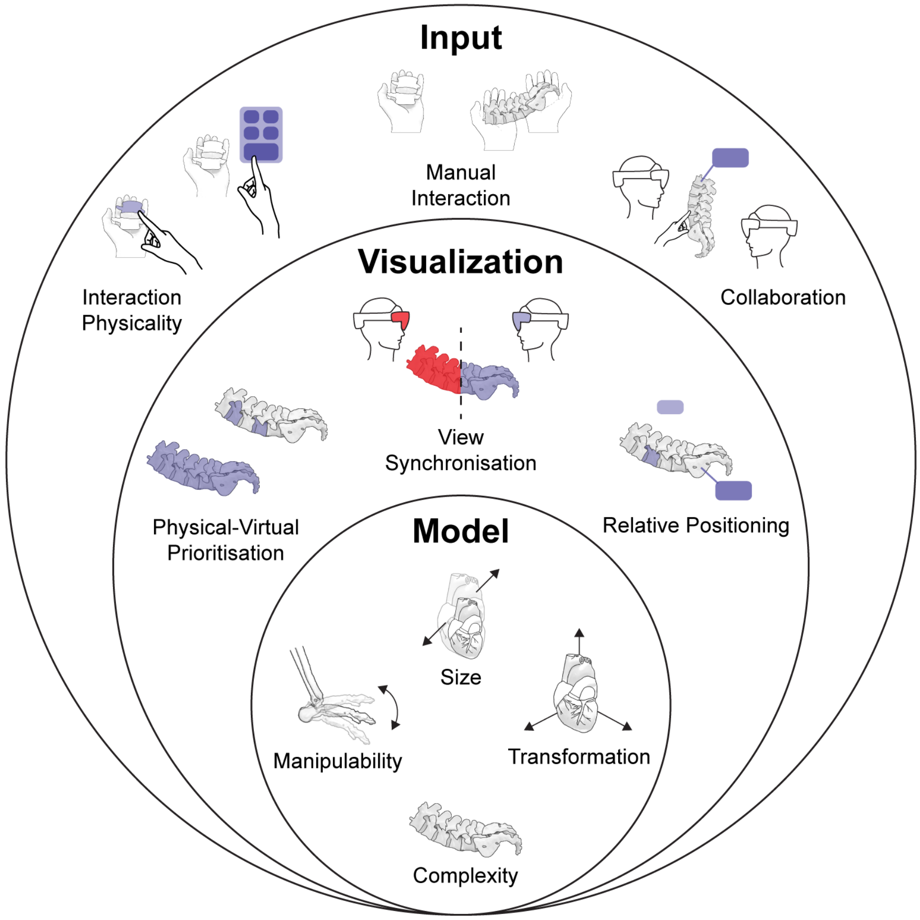


Figure 2: An overview of the design space showing a hierarchy of the dimensions

systems. By presenting the dimensions of our design space, and by demonstrating its generative power, we provide a design tool and source of inspiration for HCI researchers seeking to develop tangible AR systems that utilise physical anatomical models.

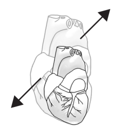
Many papers have evaluated approaches for combining AR visualisations with other physical objects, and for anatomical education in AR, however, there is no existing design space at the intersection of these concepts. We take inspiration from work by Satriadi et al. [41] who created a design space encompassing the interplay between AR and physical globes. While some dimensions in the design space by Satriadi et al. [41] are also applicable to anatomical models, we believe that anatomical models present additional distinct design dimensions that are not addressed in prior work.

We iteratively developed our design space alongside researchers in HCI and clinical biomechanics to create a first-order approximation of dimensions that pertain to the combination of AR visualisations and physical anatomical models. We categorise these dimensions into three categories: *model*, *visualisation*, and *input*, although connections between dimensions exist across the categories. We present these categories as a hierarchy, shown in Figure 2, which represents a potential tangible AR system structure. While our design space focuses on anatomical models as the underlying physical object in these visualisations, we believe many of the dimensions presented are applicable for other physical objects, including other scientific models, such as molecular or planetary models.

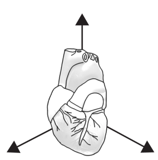
As with all design spaces, the dimensions we have presented here are not exhaustive. For our work, there is a potentially endless set of design dimensions pertaining to AR and physical anatomical models. While we are not able to address all of the factors that influence the design of these systems, our design space can be thought of as an *instance* of the wider conceptual space [5, 9, 11]. As the underlying technologies continue to develop, the design dimensions presented here will evolve and will present opportunities for researchers to modify and extend this space to incorporate new technological capabilities in AR and fabrication.

3.1 Model Dimensions

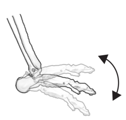
These dimensions capture the physical properties of the model pertinent to their use in AR visualisations. As the model can underpin the entire visualisation, its physical properties may determine which methods of interaction are feasible. In many cases, existing models will be used as the basis of these visualisations, meaning these models' properties will mainly be predetermined.



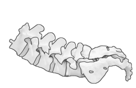
M1 Size: The range of anatomical model sizes can vary significantly, ranging from small hand-held models that are freely movable, to large immobile models such as life-size skeleton models. It is important to consider size thresholds at which models become unwieldy to manipulate using *unimanual* and *bimanual* grips, as models that require *bimanual* grip to hold comfortably will inherently limit the available interaction methods, i.e. any interactions that require manual input (see Dimension I2) will not be possible or will be severely restricted. The size of the model also impacts how much of the model can be augmented at any single point in time. For devices with a relatively limited field of view, larger models may not be able to be augmented in their entirety, simultaneously. Additionally, for particularly small models such as models of the inner ear, their size may limit the interpretability of associated AR visualisations, and may present issues in terms of object tracking reliability.



M2 Transformation: The model's number of Degrees of Freedom (DoF), ranging from 0 DoF (i.e. **static**) to 6 DoF (i.e. **freely movable** in 3D space). These DoF can be inherently limited by the size and weight of the model, or stands or other structures that restrict translation or rotation around specific axes. These constraints directly impact possible AR interaction techniques and the ease to which visualisations can be navigated.



M3 Manipulability: How the *geometric representation* of the model is able to change either through **material deformation** or through **articulation**. This could include models made from soft, deformable material (e.g. [30, 37]), or models with moveable joints. Model manipulability can be a crucial factor for integrating AR visualisations, as non-rigid structures are often more difficult to reliably track and augment. This dimension is particularly important within the context of anatomical models, as many have moving parts, or individual structures that can be manipulated.

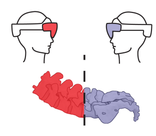


M4 Complexity: The complexity of the model. The complexity of a 3D shape can be defined using a wide range of measures. Rossignac [40] highlighted five aspects of shape complexity, including *morphological complexity*, which measures the smoothness of the shape and the size of any features, and *combinatorial complexity*, which measures the number of vertices in a polygonal mesh. Morphological complexity can affect the graspability of objects, and a higher complexity is generally seen as a benefit for tracking the model, in that tracking algorithms are more easily able to distinguish between the model and any background shapes. For the purpose of this design space, the term 'complexity' refers to *morphological complexity*, as we believe this measure of complexity has the greatest impact for AR interactions.

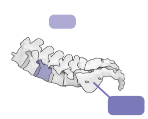
3.2 Visualisation Dimensions

These dimensions capture characteristics of how virtual elements of the visualisation are displayed to the user.

V1 Physical-Virtual Prioritisation: The extent to which real and virtual elements are prioritised in the visualisation. We define **physically-prioritised** systems as using the physical model as the primary focus of the visualisation, with virtual elements using sparingly and for the purposes of enhancing understanding of the model. Conversely, **virtually-prioritised** systems use the model purely as a display medium, with the physical form of the model encoding minimal data and primarily used to provide spatial context or tangible input.



V2 View Synchronisation: Describes the ability for multiple users to concurrently view the same virtual content. A **non-synchronised** system is one that features two or more independent devices that cannot automatically synchronise their views. A system with **synchronised** capabilities allows one or multiple users to take 'control' of the system to provide a shared view between all users.



V3 Relative Positioning: Describes the location of virtual elements in relation to the model. In their design space exploring AR and tangible globes, Satriadi et al. [41] describe several spatial relationships between virtual elements and physical globes, namely, *above* the globe, *around* the globe, *side-by-side*, and *overlaid* on the globe. We adapt this work to be more applicable to anatomical models, and the nature of the associated visualisations. We describe this dimension using five spatial relationships. The first two relationships concern virtual elements displayed within the geometric bounds of the physical model, namely, **internal** and **overlaid**:

(1) **Internal:** Virtual elements are displayed *within* the model and would support visualisations of subsurface or microanatomical structures and processes, such as internal cell growths.

(2) **Overlaid:** Virtual elements are displayed directly on, or protruding from, the *surface* of the model and could be used to visualise anatomical features such as skin, the surface of bones, and protrusions such as bone deposits.

The final three relationships concern virtual elements outside the geometric bounds of the model that may be used to communicate additional information or provide means of virtual interaction, namely, **surrounding**, **adjacent**, and **detached**:

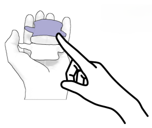
(3) **Surrounding:** Virtual elements are displayed *around* the model. These elements are displayed in close proximity to the model and may be used to highlight specific areas of interest. There include labels or other contextual UI elements.

(4) **Adjacent:** Virtual elements are displayed *next to* the model and occupy separate spaces within the visualisation, however, the positions of the virtual elements are bound to that of the model. Such a relationship could be useful for concurrent side-by-side visualisations of physical and virtual models to enable comparisons or to highlight differences.

(5) **Detached:** Virtual elements are displayed *separately* from the model and their positions are *independent* of the model. This would describe the relationship used for static UI elements, or for tangible input devices whereby the model acts as a controller for a separate virtual visualisation.

3.3 Input Dimensions

These dimensions define how users interact with both physical and virtual elements of the visualisation.

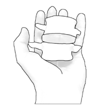


I1 Interaction Physicality. A fundamental consideration when designing interactions for AR visualisations using anatomical models is the utilisation of the model's physical properties. In other words, *to what extent are physical properties of the model are being utilised as an interaction modality?* To present this dimension, we define the **Interaction Physicality Continuum** (Figure 3). The left of the continuum represents purely physical interactions, whereby model properties such as pose, geometry and deformability are directly utilised for interactions with the system. The right represents purely virtual interaction, whereby interactions are only supported through virtual elements and do not utilise any physical properties of the model. In order to illustrate various points on the continuum, we describe three examples of interaction methods moving from the physical extremum to the virtual extremum:

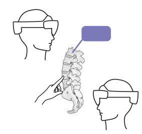
Tactile Input: These interactions utilise tactile properties of the model to provide input through methods including touch and deformation. This form of input was implemented by Murakami et al. [31] in their system *DO-IT*, in which virtual 3D shapes are deformed through the use of a deformable input tool made of polyurethane foam. While tactile input may be more difficult to achieve for pre-existing anatomical models, there is potential for the creation of bespoke 3D printed models designed to support some form of tactile input, such as supporting capacitive touch sensing using conductive filament or ink.

Geometrically Aligned: These interactions utilise virtual elements such as buttons or sliders placed around the model. The only physical property of the model that affects interaction is its geometry, which determines where the content is located relative to the model.

Object Independent: These interactions are entirely independent of the physical properties of the model, i.e., the model could be substituted for any other model without the method of input being affected. For example, this could represent interactions with a floating menu to change content views.



I2 Manual Interaction: The number of hands used for interaction. The three values for this dimension are **non-manual**, **unimanual**, and **bimanual**. These values describe the use of hands as a direct method of input to the system and does not cover the use of hands to purely support the model. For example, holding a model in view to interact with a virtual UI would be classed as **unimanual** interaction. Using two hands to rotate the model, or to pinch and scale virtual UI elements, would be classed as **bimanual** interaction. Examples of **non-manual** interaction include gaze and speech.



I3 Collaboration: Describes the number of users viewing the visualisation, and how these users interact with both the system and each other. We detail the possible collaborative scenarios in Table 1.

3.4 Generative Power

Design spaces can provide generative power for designers in a variety of ways, be it inspirational or descriptive [16]. A design space can function as a starting point for those uncovering new design challenges by presenting a set of known parameters that can be explored and combined to guide the design of the work addressing these challenges [11]. To demonstrate the generative power of our design space, we take inspiration from Eriksson et al. [11] who describes several ways of using design spaces as both generative design and re-design tools:

3.4.1 Design Through Synthesis. This method involves selecting a combination of dimensional patterns as a starting point for designing a system. For example, starting with a large model (*Size*) that can be manipulated through articulation (*Manipulability*) could represent a system utilising a life-sized skeleton model which displays information to the user about the different bones and joints through surrounding labels and descriptions (*Positional Relationship*) that are geometrically aligned (*Interaction Physicality*) with the physical model.

3.4.2 Design Through Inspiration. This method involves analysing the design of an existing system, and using the design space to categorise and describe all or a subset of its design dimensions, to act as a starting point for a future system. For example, the *AEducaAR* system [8] can be disaggregated and represented across several design dimensions. Their system uses a medium-sized (*Size*) and freely movable (*Transformation*) skull model, with AR visualisations of skull and eye anatomy overlaid (*Relative Positioning*) onto the physical skull model. Users are able to collaboratively view the visualisations, and can both control the virtual models being shown (*Collaboration*). Using these dimensions as a starting point, one could design a system which allows a teacher to teach a medical student about certain ocular conditions such as glaucoma and cataracts, using an AR-enhanced eye model.

3.4.3 Re-Design Through Enhancement. This method involves identifying missing dimensions from an existing system, and exploring

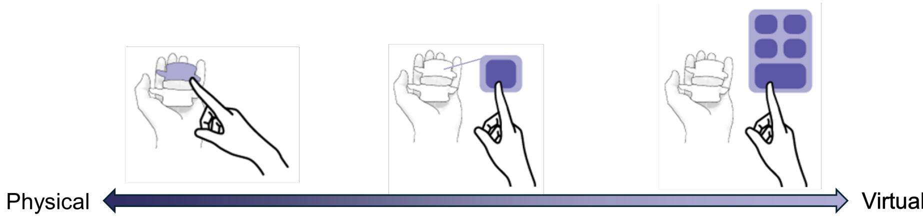


Figure 3: The Interaction Physicality Continuum

	Single-User Interaction	Multi-User Interaction
Single Viewer	No collaborative interaction.	Users who cannot see the visualisation are able to interact with the system in some form, such as through manipulating the physical model or through a separate interface.
Multiple Viewers	A single user performs the interactions while multiple users view the visualisation. Such scenarios will likely occur in learning contexts where a single user will deliver educational content to multiple viewers.	All users are able to view and control the system. It is important to consider how control over interactions is managed in order to avoid conflict between users.

Table 1: Collaborative interaction in mixed physical model/AR systems.

whether the system could be improved by incorporating these dimensions. Again, using the *AEducaAR* system [8] as a starting point, one dimension that is missing is some form of *Collaboration*. One could envision a modified version of their system which allows a teacher and a student to simultaneously and collaboratively explore the learning material to support the learning process.

4 Apparatus

In order to explore our research questions, we created an educational system for learning about spinal anatomy and axSpA disease progression based on a subset of our design space dimensions. Three versions of this system were created to allow for a comparative evaluation, comprising two AR versions (Section 4.2) and one screen-based version. Each system allowed the user to navigate a virtual 3D vertebrae model through distinct interaction techniques.

4.1 Learning Material

The learning material consisted of a series of digital 3D models and animations showing several anatomical features of the spine, and several aspects of axSpA disease progression [38] (Figure 4). In addition to the models and animations, additional educational content was provided, including labels and textual explanations of spinal anatomy and axSpA disease progression.

Healthy Spine (Figure 4A): Shows a healthy spine with no symptoms of axSpA.

Inflammation (Figure 4B): Shows inflammation of the enthesis, around the intervertebral disc and facet joints.

Fat Metaplasia (Figure 4C): Shows the formation of fat at the sites of inflammation.

Bone Erosion (Figure 4D): Shows erosion of the vertebral bone.

Syndesmophyte Formation (Figure 4E): Shows the formation of bony growths from within spinal ligaments, known as syndesmophytes.

Fused Vertebrae (Figure 4F): Shows fusion of the vertebrae, known as ankylosis.

4.2 System Designs

We created two educational AR systems as the basis of our user studies—one tangible AR system and one virtual AR system. The systems were developed in Unity, using Microsoft's Mixed Reality Toolkit (MRTK) and Vuforia [36]. Each AR system was deployed to two Microsoft Hololens 2 headsets to support collaboration. The tangible AR system utilised a physical vertebrae model as an underlying display and interaction modality, and the other system displayed the learning material as purely virtual holograms with no tangible element. We also created a screen-based version to provide a baseline comparison for the AR systems. Users could interact with a UI menu to select different elements of the learning material (see Figure 6). For the AR systems, this menu was freely movable in space, and for the screen-based system, it was fixed to the right of the screen. All three systems are shown in Figure 1.

4.2.1 Tangible AR. The tangible AR system utilised a physical vertebrae model as the basis of the visualisations (Section 4.3). This model was tracked in real time, and the virtual vertebrae models were superimposed on to the physical model *(Dimension V3)* using the Model Targets library within Vuforia [36] (Figure 6). The system was virtually-prioritised *(Dimension V1)*, with the digital models

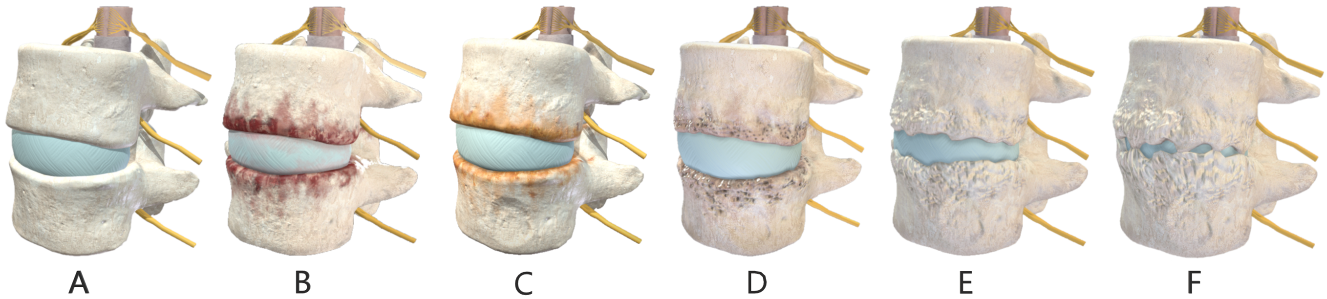


Figure 4: Digital models showing axSpA disease progression in two lumbar vertebrae. (A) Healthy Spine (B) Inflammation (C) Fat Metaplasia (D) Bone Erosion (E) Syndesmophyte Formation (F) Fused Vertebrae.

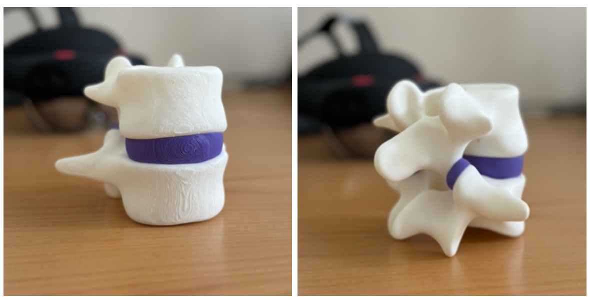


Figure 5: Front and back views of the 3D-printed vertebrae model

overlaid across the entirety of the physical model. The system allowed two users to connect simultaneously using separate headsets, and both users were able to pick up and manipulate the physical models in order to view the virtual model from any orientation *(Dimension I1)*. The superimposed virtual model matched the position and orientation of the physical model at all times, and therefore, each user's view of the virtual content depended on their position relative to the physical model.

4.2.2 Virtual AR. The virtual AR modality displayed the virtual vertebrae model positioned in world space in front of the user. The model could be scaled, and its position could be manipulated in 6DOF through direct manipulation using familiar hand-based gestures (e.g. pinching, dragging, rotating) implemented as part of the MRTK [27]. The system also allowed two users to connect simultaneously using separate headsets, with any movement or scaling of the model synced between the two users to provide a shared view.

4.2.3 Screen-Based. The screen-based system used the same underlying components as the AR systems, including the same model and menus, displayed using a desktop application developed in Unity (Figure 7). The model could be scaled and rotated using the mouse. The screen-based system was designed to replicate as much of the functionality of the AR systems as possible (i.e. 3D visualisation and 6DOF manipulation) to make any comparison as ecologically valid as possible.

4.3 Physical Vertebrae Model

The physical vertebrae model was 3D printed from the digital model of the healthy spine that was created for the learning material (Section 4.1). We define the model as small and handheld *(Dimension M1)*, measuring 11cm in length, and was printed in white and purple PLA using a Bambu Lab P1S printer. The model could not be articulated or deformed *(Dimension M3)*, however it was not fixed down and could be freely moved in space *(Dimension M2)*.

5 Comparative User Study

We first conducted a comparative user study to understand the effects of using the physical vertebrae model as part of our tangible AR system for learning about axSpA disease progression, and to compare the effectiveness of our tangible AR system to our non-tangible AR and screen-based systems for improving knowledge gain and knowledge retention.

5.1 Study Design

We used a mixed design for our study, comparing the effects of different learning modalities on anatomical understanding. The first independent variable in our study was **learning modality** (Section 5.2), which took three values: *tangible AR* (TAR), *virtual AR* (VAR), and *screen* (S). The second independent variable was **assessment time interval ('time')**, which took two values: *immediate* and *one week*, which are explained below.

Each participant took part in two sessions, spaced one week apart. In the first session, participants completed a short training activity to familiarise themselves with the individual systems they would be using. Following this, participants took part in a learning activity lasting approximately 15 minutes. The learning activity was guided by the researcher, using the models and textual explanations in the system to convey the learning material. Participants were encouraged to ask questions and to navigate the interface however they wanted to. For the AR conditions, the researcher and the participant each wore a HoloLens 2 to be able to collaboratively view and interact with the system. Immediately after the learning activity, participants completed a short assessment (*immediate*) to determine how well they had understood the material (Section 5.4). A week later, in the second session, participants were reassessed on their understanding of the material using the same assessment

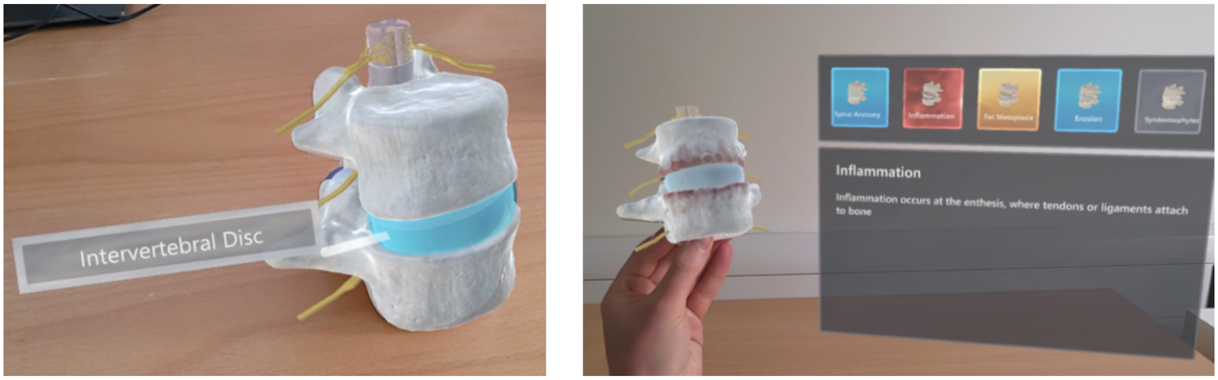


Figure 6: Views from the HoloLens 2 of the tangible AR system showing the intervertebral disc highlighted on the physical spine (left), and the floating UI menu alongside the physical spine (right).

Figure 7: The desktop system

from the first session (*delayed*). For each learning modality, the assessment was conducted using a second version of the system that included none of the learning material, such as text, animations, or highlighted areas. Participants were only able to see the virtual spine model in each case.

5.2 Independent Variable: Learning Modality

We chose *learning modality* as one of the two independent variables in our study, comparing the effectiveness of tangible AR *(TAR)*, virtual AR *(VAR)*, and screen-based (*Screen*) tools for anatomical education. Each participant used the same system for both sessions.

5.3 Independent Variable: Assessment Time Interval

The second independent variable in our study was assessment time interval ('time'), comparing assessment scores after a one-week delay. We chose one week as the reassessment delay based on previous studies assessing long-term learning retention [1, 21, 25, 39] and represents, according to the forgetting curve described by Ebbinghaus [10], a point where much of the learning material is likely to have been forgotten.

5.4 Dependent Variable: Assessment Score

Participants were assessed on their knowledge of the learning material through a verbal assessment made up of 16 short-answer questions, shown in Table 2. The questions were presented verbally after the participant was satisfied that they had understood the learning material. Participants were asked to answer the questions verbally, and their answers were assessed during the session. Participants were given a score of 1 for a correct answer, and a score of 0.5 for a partially correct answer (e.g. correctly locating inflammation around the intervertebral disc, but not around the facet joints). The questions were designed to assess both *spatial* understanding (e.g. locations of disease progression), and *non-spatial* understanding (e.g. descriptions of anatomical features and symptoms). The questions were split into two categories representing *spatial* and *non-spatial* questions, and are labelled as such in Table 2. This choice was made to understand how the spatial learning benefits of AR [34, 51] apply to the inherent spatial component of understanding anatomical structures and processes.

5.5 Additional Measures

5.5.1 Prior Experience Questionnaire. Participants were asked to indicate their prior experience with AR, their prior experience with analysing anatomical visualisations, and their prior knowledge of rheumatic conditions using 7-point Likert scales. For the prior experience questions, the scale values were: (1) Never, (2) Once or twice in total, (3) Yearly or less, (4) Every 6 months or less, (5) Monthly or less, (6) Weekly or less, and (7) Multiple times per week. For the prior knowledge of rheumatic conditions question, the scale values were: (1) Very poor, (2) Poor, (3) Somewhat poor, (4) Fair, (5) Somewhat good, (6) Good, and (7) Very good.

5.5.2 User Experience. At the end of the first session, participants were asked to fill in a 7-point Likert questionnaire (1 = strongly disagree, 7 = strongly agree) about their experience with the system and content. A list of the questions is shown in Table 3.

5.5.3 Verbal Feedback. After the first session had concluded, participants were asked if they would like to provide any additional verbal feedback about the study, system, or learning material.

5.6 Participants

A 2024 systematic review and meta-analysis of virtual reality and augmented reality in anatomy education [14] reported summary effect sizes for 27 studies investigating "XR technologies compared

ID	Question	Category
Q1	Where is the vertebral body?	Spatial
Q2	What is the purpose of the vertebral body?	Non-Spatial
Q3	Where are the facet joints?	Spatial
Q4	What is the purpose of the facet joints?	Non-Spatial
Q5	Where are the spinous processes?	Spatial
Q6	What is the purpose of the spinous processes?	Non-Spatial
Q7	Where are the transverse processes?	Spatial
Q8	What is the purpose of the transverse processes?	Non-Spatial
Q9	Where is the intervertebral disc?	Spatial
Q10	What is the purpose of the intervertebral disc?	Non-Spatial
Q11	Where does inflammation occur in the spine?	Spatial
Q12	What type of lesions can replace areas of inflammation over time?	Non-Spatial
Q13	What type of tissue are the lesions thought to be?	Non-Spatial
Q14	Where does bone erosion take place?	Spatial
Q15	Where does new bone growth take place?	Spatial
Q16	What can further bone growth lead to?	Non-Spatial

Table 2: Assessment Questions

ID	Item
Q1	I found the learning material easy to understand
Q2	I found the 3D models and animations easy to interpret
Q3	I feel like I have a better understanding of axial spondyloarthritis
Q4	I found it easy to understand explanations from the researcher
Q5	I could easily see and refer to specific parts of the visualisation
Q6	I found the overall experience engaging

Table 3: Post-Study Questionnaire Items

to traditional resources for improving knowledge after interventions", and found a medium effect (Cohen's d = 0.40) in favour of XR interventions with low heterogeneity (I^2 = 17.1%, Q = 39.7). Adjusting for publication bias gave an adjusted effect of 0.45. Based on these findings, we conducted an a priori power analysis using G*Power 3.1.9.7 [13] to determine a suitable sample size, using an effect size of 0.45, a confidence interval of 0.95, and a power of 0.80 for a two-way mixed ANOVA. This resulted in a sample size of 13 participants for each of the three conditions. In total, 39 participants were recruited for the study. Of these participants, 21 identified as female, 17 identified as male, and 1 participant identified as non-binary. Three participants chose not to disclose their age. The remaining 36 participants were aged between 19 and 59 years old ($M = 30.4, SD = 11.6$). At the start of the first session, participants rated their prior experience with using AR, their prior experience with analysing anatomical visualisations, and their knowledge of rheumatic conditions (reported in 'Results'). Participants included undergraduate students, postgraduate students, and staff working at a university in the UK. In the inclusion criteria, participants were required to have no prior knowledge of axSpA in order to be eligible for the study.

5.7 Statistical Analysis

Statistical analysis was performed using IBM SPSS Statistics version 29.0.1.1 (244) [18]. Our study design resulted in two main effects of *learning modality* and *time*, and one interaction effect between them. Normality and homogeneity of variance were tested for each combination of *learning modality* and *time* using the Shapiro-Wilk test and Levene's test respectively. Mauchly's test of sphericity was used to test equality of variances between both *time* points for each *learning modality*. A mixed ANOVA was performed to understand if there was any interaction between the *learning modality* and *time* variables on participants' assessment scores. Additionally, a one-way ANOVA was performed to compare assessment scores across each condition at both assessment time intervals.

5.8 Results

5.8.1 Participant Demographics. On average, participants reported little prior experience with AR ($M = 2.10, SD = 1.021$), little prior experience with analysing anatomical visualisations, ($M = 1.33, SD = 0.772$) and little knowledge of rheumatic conditions ($M = 1.56, SD = 0.882$). For each of the participant demographic questionnaire items, we conducted two-way ANOVAs to identify any interactions between each measure and condition on participants' combined scores across both assessments. No statistically significant interactions were found for *prior experience with AR* ($F = 1.666, p = 0.169$) , *prior experience analysing anatomical visualisations* ($F = 1.298, p = 0.287$) , or *knowledge of rheumatic conditions* ($F = 0.995, p = 0.439$).

5.8.2 Assessment Scores. Participants generally scored well on average across all three learning modalities, as shown in Figure 8. For the *immediate* assessment, the average scores were: Screen = 14.1 (SD = 1.53), TAR = 12.7 (SD = 3.09) and VAR = 12.7 (SD = 2.90). For the *delayed* assessment, the average scores were: Screen = 13.3 (SD = 2.59), TAR = 12.3 (SD = 2.85) and VAR = 12.7 (SD = 2.80). A mixed ANOVA found no statistically significant interaction between the *learning modality* and *time* variables on participants' assessment score ($F = 0.826, p = 0.446$). In addition, we found no statistically significant main effects for *learning modality* ($F = 0.794, p = 0.460$) or *time* ($F = 2.811, p = 0.102$). A one-way ANOVA also found no statistically significant effect of *learning modality* when comparing scores across *immediate* ($F = 1.227, p = 0.305$) and *delayed* ($F = 0.409, p = 0.667$) assessment intervals.

On average, participants' assessment scores decreased between the *immediate* and *delayed* assessment intervals across the different *learning modalities* (*Screen* ($M = -0.81, SD = 1.69$), *TAR* ($M = -0.39, SD = 1.87$), and *VAR* ($M = -0.04, SD = 0.80$)). This is shown in Figure 9. As described above, this was not statistically significant.

We also analysed assessment scores grouped by **spatial** and **non-spatial** questions. For the **spatial** questions, we found no statistically significant interaction between *learning modality* and *time* ($F = 0.556, p = 0.578$), and no statistically significant main effects for *learning modality* ($F = 0.478, p = 0.624$) or *time* ($F = 0.418, p = 0.522$). Similarly, for the **non-spatial questions**, we found no statistically significant interaction between *learning modality* and *time* ($F = 0.421, p = 0.660$), and no statistically significant main effects for *learning modality* ($F = 0.995, p = 0.380$) or *time* ($F = 3.184, p = 0.083$). A further one-way ANOVA also found no statistically significant effect of *learning modality* when comparing scores across *immediate* ($F = 1.227, p = 0.305$) and *delayed* ($F = 0.409, p = 0.667$) assessment intervals.

5.8.3 Post-Study Questionnaire. On average, participants rated all elements of the post-study questionnaire highly. However, there was no significant difference in rating across the three *learning modalities* ($p > 0.05$ for each question). This data is summarised in Figure 10.

5.8.4 Verbal Feedback. This section presents quotes taken from the verbal feedback collected at the end of each study session, focusing on feedback relating to the physical spine model. Quotes are presented alongside the associated participant number within each learning modality (e.g. **P1** $_{\mathbf{TAR}}$).

Most participants who used the physical spine model highlighted benefits for both understanding and interaction related to its tangibility:

> **P6**: *By having something to hold, there is value in that… it adds an extra degree of contextual information, you've the feedback, you can kind of get a feel for what it actually is.*

Additionally, **P3** $_{\mathbf{TAR}}$ described how it was *"easier to actually interact with"* the physical model, and that it was *"easier to spin it around"*. Another participant discussed their preference for tactility to support their individual learning:

> **P8** $_{\mathbf{TAR}}$: *I think I liked that I had something to play around with because I'm quite a hands-on person. So whenever I study as well, it just helps if I can look at something as well as kind of feel it and just have the freedom to look at different angles.*

Two participants highlighted how the physical nature of the model led to an increased feeling of realism: **P13** $_{\mathbf{TAR}}$ thought that *"it was quite nice to have one that I could actually touch, because it felt more tangible and real"*. **P5** $_{\mathbf{TAR}}$ also described how the physical model helped them to form a closer connection between the learning material and their own spine:

> **P5** $_{\mathbf{TAR}}$: *Having it there helps it feel like an actual part of the spine, so I sort of imagine where it would be in my body, whereas if it was just floating, it wouldn't feel quite as real.*

Although feedback was generally positive, some participants were unsure about the net benefits of including a physical model in the AR system. This was partly caused by some frustration regarding the tracking of the spine:

> **P6** $_{\mathbf{TAR}}$: *I think if the tracking was better on this, it would justify it more. With the issues of tracking it kind of almost becomes a frustration.*

In the virtual AR condition, one participant suggested the idea of a physical spine model for improving the manipulation of the visualisations:

> **P5** $_{\mathbf{VAR}}$: *Spinning it around and everything… it kind of hurt my wrist a bit… I think it would probably be nicer if I could just grab it, look at it, and then just hold it. I guess kind of like a real model of a spine.*

Another participant using the virtual AR system commented that the positions of holograms were difficult to perceive as a first-time user, and that having items in the background such as our study apparatus was disruptive. They found the lack of a blank canvas over which to visualise the holograms disrupted their perception of *"where you think it should be"* (**P10** $_{\mathbf{VAR}}$). Having a physical point of reference in the world on which to anchor the holograms may help alleviate this problem somewhat.

5.9 Discussion: Comparative Study

Given the limited and conflicting nature of prior research into the use of of AR for anatomical education, it is important that we scope the contributions of our work clearly within this space. Notably, our results showed no significant differences in knowledge gains or knowledge retention between the three conditions. This aligns with prior work by Cercenelli et al. [8] who also found no significant difference in learning outcomes between their tangible AR system and a standard human anatomy atlas for learning skull and eye

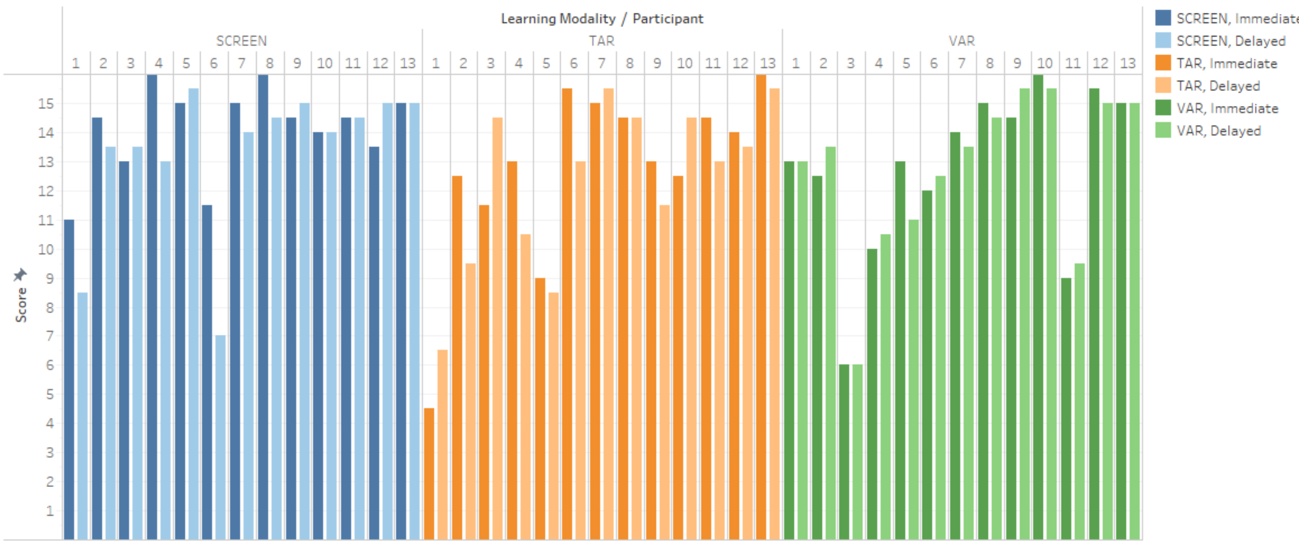


Figure 8: Participants' assessment scores

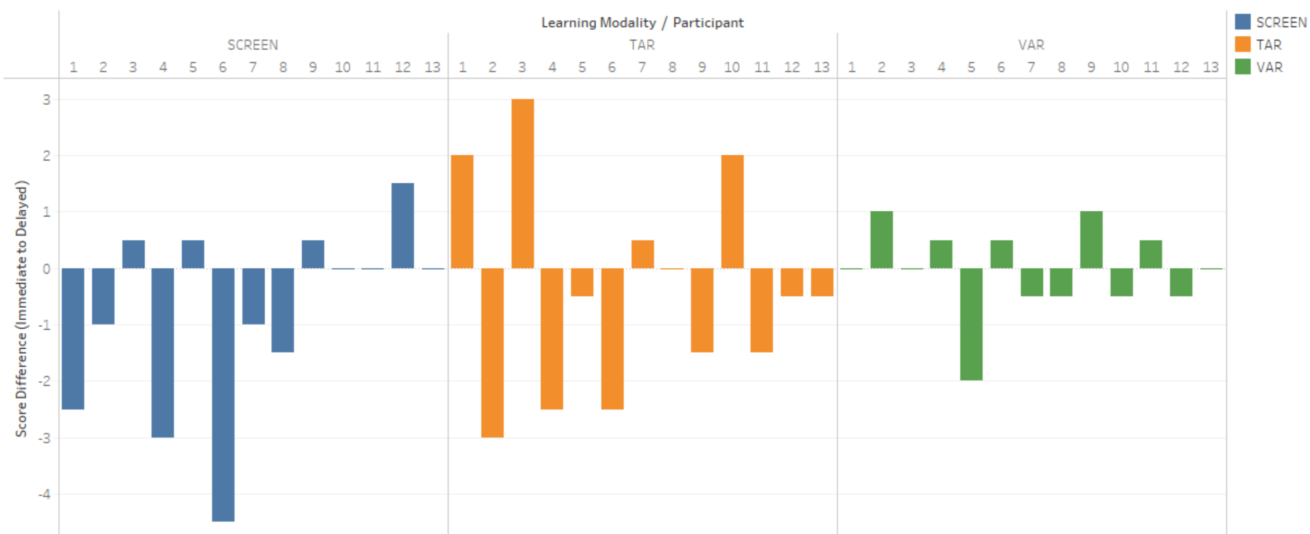


Figure 9: Change in scores between first and second assessments

anatomy. This also aligns with previous work outside of anatomical education by Knierim et al. [22], who found that underlying physical interfaces had no effect on comprehension or knowledge transfer for learning physics in AR. While García-Robles et al. [14] report improved knowledge gains for AR interventions in their systematic literature review, it is worth noting that the majority of control conditions in the included studies utilised 2D textbooks, atlases, and lectures. While these control conditions may represent what is traditionally used for teaching anatomy in many pedagogical scenarios, it is unclear as to whether the reported improvements in knowledge gains are as a result of participants using *any* interactive 3D system, or whether these improvements are related specifically to the benefits provided by AR. Our study sought to address this lack of clarity in two ways. Firstly, we compared a tangible AR system with a non-tangible AR system to isolate any effects of introducing a physical model. Secondly, we compared both AR systems with a non-AR equivalent running on a desktop PC, which we designed to replicate as many of the affordances of AR as possible to provide the most meaningful comparisons. Crucially, the desktop version provided users with a 3D model, which they could freely manipulate in 3D space using the mouse.

Overall, there was very little loss in knowledge over the one-week 'forgetting' period across all three conditions (Section 5.8.2). Interestingly, our results do not align with the forgetting curve proposed by Ebbinghaus [10], however, previous research has shown

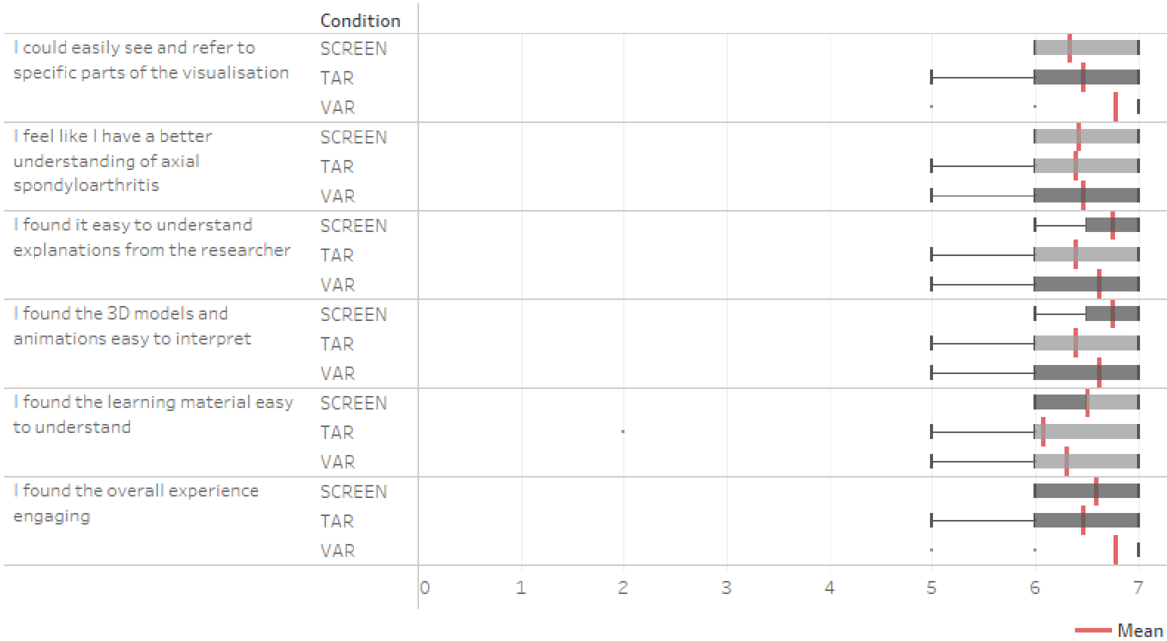


Figure 10: Post-study Likert questionnaire responses

that AR provides significant benefits to knowledge retention compared to non-AR learning modalities [19, 43], and prior work has observed this contradiction of the forgetting curve for AR systems [24, 50]. Based on this prior work, it is likely that the highly engaging nature of the AR visualisations contributed to this contradiction of the forgetting curve in our study. Other factors that may have contributed include the difficulty and complexity of the learning material, and the high proportion of university students and staff as participants.

While we did not observe any significant differences in knowledge improvement or retention between the different learning modalities, the interpretation of our findings should consider the concept of "transfer-appropriate processing" [29]. This concept suggests that "the value of any particular acquisition or practice condition can only be evaluated when considered in the context of the particular transfer test used to evaluate learning" [23]. Participants in our study were assessed using the same modality they used for learning session. The choice to use the same system for the assessments was done to minimise any uncertainty introduced by requiring participants to map and transfer their knowledge from one medium to another. However, this choice also meant that we were unable to measure the effect of transfer-appropriate processing on learning outcomes. This presents an opportunity for future work to investigate how the use of physical models affects knowledge gains and retention when assessed through a modality different to that of the learning material. Additionally, as the three learning modalities were comparable in terms of their three-dimensionality, interactivity in 6DOF, and identical learning material, this helped to reduce any biases introduced by using the same learning modality for both learning and reassessment.

From a learning experience perspective, all three systems were largely seen as useful and engaging. Participants across each condition found the ability to visualise and manipulate a 3D spine model beneficial for learning, and the novelty of both AR conditions provided an enjoyable learning experience for many participants. Although the physical model introduced some difficulties regarding tracking, the overall effect was generally seen as positive, particularly for enabling more intuitive manipulation of the visualisations and for utilising the sense of touch to aid learning. The net benefits of incorporating a physical model are still unclear, and this will largely depend on the individual user in terms of their preferred learning style and confidence using the system. As both AR and object tracking technologies improve, it is likely that difficulties surrounding the tracking of physical models will have less of an impact on the learning experience, and that this will make tangible AR systems more accessible and useful for a greater number of users.

Although the effectiveness of tangible AR over virtual AR for anatomical education requires further investigation, the benefits highlighted by participants around tangible interaction and identifying with the physical model present unique opportunities for patient education. The ability for patients to form a more personal connection with the learning material through a physical representation of their own body, combined with contextualised holographic overlays, may help to provide patients with a deeper understanding of their conditions, and could ultimately lead to positive self-management outcomes.

6 Expert Evaluation with Clinicians

To explore opportunities and challenges around the use of tangible AR for anatomical education in clinical practice, we conducted a qualitative evaluation of the system with clinicians working in axSpA care in the Royal United Hospitals, Bath. We invited three clinicians working in different roles within the axSpA care pathway to take part in a demonstration and exploration of the system, followed by a semi-structured interview to gather their thoughts on the system and potential implications for using such systems in clinical practice.

ID	Clinical Role
C1	Clinical Nurse Specialist
C2	Clinical Physiotherapist
C3	Consultant Rheumatologist

Table 4: Details of Participating Clinicians

6.1 Participants

We recruited participants using an expert sampling method, ensuring a broad range of clinical roles to explore as many areas of the patient care pathway as possible. In total, three clinicians were recruited through word-of-mouth at the Royal United Hospitals, Bath, summarised in Table 4.

6.2 Study Process

Participants was asked to attend an individual in-person session lasting approximately 60 minutes each. In each session, the clinician was given a short period to familiarise themselves with the system and the HoloLens 2, before taking part in a guided exploration of the system lasting approximately 20 minutes. Following this exploration, one-to-one semi-structured interviews were conducted, lasting approximately 30 minutes. The questions focused on current educational practices, the design and usability of the AR systems, and opportunities and challenges for the use of in clinical practice. A full list of the interview questions are included in Table 5. Audio recordings were taken during the interviews, which were then transcribed and fully anonymised by the researcher. Thematic analysis was performed on the resulting transcripts with a general inductive approach [45], using NVivo qualitative data analysis software [26].

6.3 Results

By analysing the interview transcripts, we identified several themes: educational benefits of AR, educational benefits of the physical model, system usability, and barriers to clinical integration. This section details these themes, substantiated by quotes taken from the transcripts. We have indicated which participants contributed to individual themes at the start of each subsection. While the interviews were transcribed verbatim, some the quotes we present have been edited purely for the purpose of removing unintentionally repeated words and filler words such as "umm" and "uhh". The meaning and context of each quote was carefully preserved.

6.3.1 Educational Benefits of AR. (**C1, C2, C3**)

The educational benefits to patients of the AR system were discussed by all of the clinicians. In particular, the active participation and interaction involved with the system was seen as beneficial.

> **C1**: *The fact that you have to reach in to move things and press the buttons, it feels like you're engaging with it - it's not just a passive experience.*

> **C2**: "*But the interaction with it, really interesting. I found it really interesting. . . . I really like being able to turn the model around, I like being able to zoom in a little bit.*"

> **C3**: "*I really like [the system]. I think it's a really good way of visually demonstrating to patients what's happening and I think obviously the the benefit of doing it in augmented reality is it allows a little bit of manipulation so that the patient can kind of see a 3D image from all planes rather than just a picture.*"

In addition to patient education, **C2** also commented that the system would "*probably help a lot of clinicians as much as patients.*" They discuss the fact that even experienced clinicians sometimes find it hard to interpret and orientate themselves around MRI images, and that "*an overlay of these models could be quite useful for clinicians who don't look at scans that often.*"

6.3.2 Educational Benefits of the Physical Model. (**C1, C3**)

Both **C1** and **C3** expressed a preference for the inclusion of the physical vertebrae model, and commented on the educational benefits of its tangibility.

> **C1**: *I quite like [the physical model] though, because it's quite nice being able to pick something up that's got some substance to it and then you've got all the details superimposed . . . I think that's really quite helpful, been able to actually hold something and feel it. Yeah, that's really useful*

> **C3**: "*I liked having the physical spine. I think that that works really well . . . I think it does aid understanding if you've got that extra sense involved in it, that it's not just visual, you've got that tactile involvement as well.*"

Another benefit identified by **C1** was that of self-association, i.e., being able to relate the AR models with their own body.

> **C1**: *You see something like this or you see that 3D model. . . I can look at that and think I know a bit of my body looks like this. Whereas if you look at a booklet or something playing on the screen, yeah, it's just quite abstract.*

6.3.3 System Usability. (**C1, C2, C3**)

All three clinicians commented on the usability of the AR system, including from a personal perspective, and from the perspective of potential patient use. **C2** and **C3** reported some difficulties regarding interacting with the holographic elements of the system during the guided exploration.

ID	Question
Q1	What does a typical session with patients involve?
Q2	How do you educate patients about axSpA?
Q3	What are your first impressions of the system?
Q4	Could you see a system such as this being a part of your sessions with patients?
Q5	Are there any features of the system you particularly like?
Q6	Are there any features you would change?
Q7	What benefits do you think AR could provide in practice compared to traditional methods?
Q8	What challenges do you envisage for the use of AR in clinical practice?
Q9	Are there any aspects of axSpA education in particular that you think would not be suitable for AR-based learning?
Q10	Is there anything else you wish to mention that hasn't been discussed already?

Table 5: Interview Questions in the Expert Qualitative Evaluation

> **C2**: *I have little experience with VR, AR . . . It felt quite clumsy with it not being something I'm particularly used to. Whether other people are more intuitive with how to grasp and turn, I don't know.*

> **C3**: "*I personally struggled with interacting with the menu a bit and it was just sort of getting used to how to have that stabilized in the right location for me to be able to look at it was something that took a little bit of practice.*"

Regarding the visual clarity of the system, **C2** also highlighted challenges with focusing on holographic elements in a visually noisy environment:

> **C2**: *Practically, I think that one thing I quite struggled with was the ability to see through, then the ability to sort of disassociate from the background. I think that's what I found almost a little bit, not quite headache inducing, but tiring for my eyes, is being able to - the stuff on the back wall behind - to kind of ignore that and try and just see the model and the menu option.*

Conversely, **C1** reported that the system was "pleasant to use" and that their first impressions were "really positive". However, they also noted an initial lack of awareness of the UI elements that were positioned further below the eyeline. In part, this was likely a result of the limited vertical FOV provided by the HoloLens 2.

C1 and **C3** also highlighted potential usability implications for people with restricted mobility, such as those with peripheral symptoms of axSpA including stiffness and difficulties with motor movements in their hands. **C3** talked about the weight of the HoloLens 2 itself, commenting that "*. . . it's not heavy, but it's not a negligible amount of weight for somebody who's already got a lot of stiffness in their neck to be carrying.*" **C3** goes on to mention the importance of "*. . . making sure that the patients are made comfortable and well supported around how to wear the headgear and maybe some advice on some stretches and things they can do afterwards as well.*"

Regarding interacting with the AR elements of the system, **C1** suggested that while some patients experience peripheral symptoms in their hands, these are quite often related to strength and fine motor movements, affecting tasks such as holding pens or doing up buttons. For our AR systems, **C1** suggested that many of the interactions such as moving the holograms and pressing UI buttons "*. . . might work quite well for them, so that might not be too much of a challenge, really.*" **C1** also highlighted how the accessibility of the system could be beneficial for patients with restricted mobility who enjoy engaging with novel technology.

> **C1**: *. . . they might feel quite excited by the fact that it is something they can use and they're not struggling with . . . it's nice to know that people have got things that they can engage with . . .*

6.3.4 *Barriers to Clinical Integration.* (C1, C2, C3)

One of the main barriers for integrating new activities within clinical workflows is limited patient-facing time. **C1** emphasised that "*in clinic every second counts, literally every second. . . from the minute you call them from that waiting room, to how long they take to walk to the [room]*". All three clinicians saw potential value in the AR systems when used as part of separate or group learning sessions.

> **C2**: *If we had 7 or 8 headsets in a room, 6 or 7 people with one clinician. . . instead of having the the PowerPoint slides up, we'd have a model and we'd been moving the model and rotating the model and prefacing what the model is. . . It would be interesting to see what people thought about it, how they engaged with it if we had the opportunity to do that.*

> **C3**: "*. . . it's definitely something I could see being really helpful in something like a group education session, alongside maybe a guided talk or something to help patients understand what we're talking about.*"

One challenge that all clinicians highlighted was the importance of user confidence in using the system effectively. **C1** and **C3** discussed this from the patient perspective:

> **C3**: *I think that if we were gonna use these, it would need some- probably a human being, to be present to help the patient overcome some of the technical aspects of that. So you were giving me quite a lot of coaching*

> *there on how to get the most out of the system. And I suspect that wouldn't be notably different for patients.*

When discussing the use of a similar system to encourage patient engagement, **C1** commented that especially for those who are not overly tech savvy or interested, *"If we just said to them, look, here we've got this headset, this is what you do, I'm just gonna get it going for you, do you wanna have a little go? ... then they'd feel quite supported whereas if you send people away they're not always quite going to look at it or they might be a bit daunted by accessing it."*

Additionally, **C2** highlighted clinician confidence when asked about practical challenges for implementing AR systems in day-to-day clinic activities. Specifically, they mentioned the importance of *"having clinicians confident enough like you were to describe what I'm seeing and you're not seeing."* This refers to the ability for clinicians to feel confident enough to explain the AR visualisations to patients in real time without necessarily being able to see the content themselves.

6.4 Discussion: Expert Evaluation

One of the primary goals of patient education is to provide patients with the information they need to make informed decisions about their health. As with many chronic conditions, these decisions often take place over long periods of time, often over the course of a lifetime. For axSpA specifically, consistent long-term self-management practices, such as stretching and exercise, are crucial in maintaining a good quality of life for patients. Education around their benefits is important for motivating adherence to these practices, and patients who have a greater understanding of their conditions tend to have better outcomes.

Despite its potential, our results highlight significant challenges for the integration of tangible AR tools within clinical workflows. One of the main challenges lies in providing patients with access to the AR systems and learning material. As highlighted by the clinicians in our study, the use of such systems would likely require in-person supervision to ensure both ease of use and contextualisation of learning material. Like in many healthcare contexts, clinical workflows in axSpA are tightly constrained by the time available for each patient, and the integration of novel technologies such as AR should aim to complement existing workflows, rather than compound the issues within them [17]. One possible solution presented by clinicians was to centralise access to the AR systems, and deliver the content to multiple patients in a single, collaborative session.

For conditions such as axSpA that can restrict physical mobility, head-mounted AR systems such as the HoloLens 2, and certain interactions used within them, can present barriers for patients in terms of accessibility and usability. One clinician in our study highlighted an example of this, in which a patient with stiffness in their neck may struggle to wear a headset for extended periods of time. Another example of this would be a patient with peripheral symptoms in their hands or fingers, who may find it difficult to perform gestural interactions such as pinching and wrist rotation. Possible solutions to these problems could involve patients taking a more passive role regarding interactions, allowing clinicians to control a shared view of the learning material; and providing non-AR alternatives to limit any educational inequalities for patients who are unable to use the AR systems.

7 Discussion

7.1 Inter-Study Comparisons

As highlighted through our studies, there is clear potential for both tangible and non-tangible AR to provide educational benefits for patients within axSpA care. Across both studies, we found notable similarities in participants' impressions of our systems, which provide insights for the design and implementation of future systems:

7.1.1 Benefits of the Physical Model. Our results show that both lay-users and clinicians saw educational value in the physical spine model over the purely holographic equivalent. Participants in both studies highlighted the cognitive benefits of the model's tactility, and several participants in the comparative evaluation reported that the model enabled easy and intuitive interaction with the virtual content.

Another benefit highlighted across both studies was that of realism and self-association with the physical model, i.e., the idea that the physical model represents part of the user's own body. For lay-users, this connection has the potential to improve the overall understanding of anatomy in relation to the body as a whole. For patients, this may support a deeper understanding of their own conditions, which may ultimately contribute to improved health outcomes. This concept of self-association could be taken further in future work by using volumetric patient CT data to form the basis of the virtual content and the underlying physical model, to provide patients with personalised visualisations of their own anatomy.

7.1.2 System Usability and Accessibility. Participants in both studies highlighted some limitations of the tangible AR system regarding the tracking of the physical model, and some participants in the user study found this inconsistency in tracking resulted in increased frustration. However, this was not the case for most participants, and as the underlying technologies improve, system performance should present less of a barrier for users in interacting with the tangible AR systems more efficiently.

Our results also highlight challenges in terms of accessibility. One participant in the user study described how the action of rotating the holographic spine in the virtual AR system led to some wrist pain, and suggested that they would benefit from holding a physical model to manipulate the visualisations more intuitively. One of the clinicians commented that for patients with restricted mobility or stiffness, as is fairly common in axSpA, the weight of head-mounted devices such as the Hololens 2, as well as the gestures required to interact with our systems, could present challenges.

7.2 Limitations and Future Work

The main limitation of our work comes from its narrow clinical scope. Contributing factors for this include our focus on one form of spinal arthritis, and our expert evaluation involving a small number of clinicians within a single hospital in the UK. Given the number of design possibilities, our tangible AR system has only scratched the surface of what is possible in this space, and there may be potential for other design choices to highlight additional educational benefits.

Future work could explore other facets of tangible interaction, such as material deformability or capacitive touch input.

Secondly, although participants were required to have no prior knowledge of axSpA, reported little prior experience with analysing anatomical visualisations, and little knowledge of rheumatic conditions, the lack of an explicit pre-study baseline assessment limits the conclusions that can be drawn regarding knowledge gains.

Finally, our evaluation with non-patient participants limits the insights our work can provide into the benefits of tangible AR systems for improving patient outcomes. There may be implications specific to patient education that we were not able to observe in our studies, such as the impact of personal connections with the learning material on educational outcomes, which could be explored in future research. Although our choice of a one-week reassessment delay does not represent the long-term for people living with axSpA and other chronic conditions, this provides the basis for future work to investigate the lasting effects of educational interventions such as our AR systems for motivating adherence to self-management practices, and ultimately for improving patient outcomes.

8 Conclusion

While existing work has shown numerous advantages of AR for anatomical education, the cognitive benefits afforded by traditional physical learning modalities, such as anatomical models, are not retained in most AR systems. The purpose of our work was to explore the combination of physical models and AR visualisations to support anatomical education. Specifically, we wanted to understand how the affordances of physical anatomical models and AR can be utilised to create tangible AR systems (RQ1), the learning benefits that physical models can provide over equivalent non-tangible AR systems (RQ2, RQ3), and opportunities and challenges around the use of physical models and AR for anatomical education in clinical practice (RQ4).

To answer our research questions, we first presented a design space encompassing the interplay between physical anatomical models and AR. We demonstrated its generative power through several examples of system design and re-design processes, showing how it can be used as an inspirational design tool for creating future systems (RQ1). Based on our design space, we created a tangible AR system for learning spinal anatomy and axSpA disease progression using a physical vertebrae model, along with equivalent non-tangible AR and screen-based systems. In order to assess the learning benefits of the physical model, we compared the effectiveness of these systems for improving knowledge gains (RQ2), knowledge retention (RQ2), and learning experience (RQ3) over a one-week longitudinal study with lay-users. Our results showed no significant difference in knowledge gains or retention between the three systems, however participants' learning experience was generally positive when interacting with the physical model. In particular, participants valued the realism and self-association provided by the physical model, as well as its benefits for intuitive interaction and manipulation of the visualisations. We also conducted a qualitative evaluation of our AR systems with clinicians working in axSpA care through a guided exploration of the system and semi-structured interviews (RQ4). These highlighted several implications for the use of both tangible and non-tangible AR for patient education in clinical practice, including challenges around system usability and user confidence.

This paper offers several key takeaways. First, while we did not find any significant improvement in knowledge gains or retention through use inclusion of the physical vertebrae model, our results show that the realism and ease of interaction provided by physical models can benefit the overall learning experience for users. Second, based on feedback from participants, physical models have the potential to provide benefits for patients in understanding their own anatomy and disease progression through increased self-association. This could allow patients to form a deeper connection with the learning material, and may ultimately contribute to improved health outcomes. Finally, while the results of both studies highlight the benefits of physical models, challenges remain for implementing tangible AR systems within clinical practice. We need to understand how such systems can be effectively integrated into time-constrained clinical workflows and how these systems can be designed to be accessible to wider patient populations.

Our work highlights many opportunities for further research into tangible AR systems for anatomical education, including investigating its benefits for patients in understanding their own anatomy and disease progression, and exploring how other attributes and affordances of physical models can be utilised to create novel forms of learning anatomy through tangible AR. As the increasing maturity and prevalence of AR technologies continue to support the delivery of anatomical education, the incorporation of physical models holds significant potential to provide more realistic, engaging, and intuitive learning experiences.

Acknowledgments

We would like to thank Imogen Ridley [38] for creating the 3D models and animations used in our AR systems. We would also like to thank Kim Sauvé for helping to create the images used in our design space. This work was partially funded by the Engineering and Physical Sciences Research Council (reference 2441867) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (FORCE-UI, Grant agreement No. 853063).

References

[1] Mariam Baghdady, Heather Carnahan, Ernest W N Lam, and Nicole N Woods. 2014. Test-enhanced learning and its effect on comprehension and diagnostic accuracy. *Medical education* 48, 2 (2014), 181–188. doi:10.1111/medu.12302

[2] Florence Bara, Edouard Gentaz, Pascale Colé, and Liliane Sprenger-Charolles. 2004. The visuo-haptic and haptic exploration of letters increases the kindergarten-children's understanding of the alphabetic principle. *Cognitive development* 19, 3 (2004), 433–449.

[3] Roghayeh Barmaki, Kevin Yu, Rebecca Pearlman, Richard Shingles, Felix Bork, Greg M. Osgood, and Nassir Navab. 2019. Enhancement of Anatomical Education Using Augmented Reality: An Empirical Study of Body Painting. *Anatomical sciences education* 12, 6 (2019), 599–609. doi:10.1002/ase.1858

[4] Mark Billinghurst, Hirokazu Kato, Ivan Poupyrev, et al. 2008. Tangible augmented reality. *Acm siggraph asia* 7, 2 (2008), 1–10.

[5] Michael Mose Biskjaer, Peter Dalsgaard, and Kim Halskov. 2014. A constraint-based understanding of design spaces. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 453–462. doi:10.1145/2598510.2598533

[6] Felix Bork, Leonard Stratmann, Stefan Enssle, Ulrich Eck, Nassir Navab, Jens Waschke, and Daniela Kugelmann. 2019. The Benefits of an Augmented Reality Magic Mirror System for Integrated Radiology Teaching in Gross Anatomy. *Anatomical sciences education* 12, 6 (2019), 585–598. doi:10.1002/ase.1864

[7] Kerem A. Bölek, Guido De Jong, and Dylan Henssen. 2021. The effectiveness of the use of augmented reality in anatomy education: a systematic review and meta-analysis. *Scientific reports* 11, 1 (2021), 15292–15292. doi:10.1038/s41598-021-94721-4

[8] Laura Cercenelli, Alessia De Stefano, Anna Maria Billi, Alessandra Ruggeri, Emanuela Marcelli, Claudio Marchetti, Lucia Manzoli, Stefano Ratti, and Giovanni Badiali. 2022. AEducaAR , Anatomical Education in Augmented Reality: A Pilot Experience of an Innovative Educational Tool Combining AR Technology and 3D Printing. *International journal of environmental research and public health* 19, 3 (2022), 1024. doi:10.3390/ijerph19031024

[9] Graham Dove, Nicolai Brodersen Hansen, and Kim Halskov. 2016. An Argument For Design Space Reflection. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (Gothenburg, Sweden) *(NordiCHI '16)*. Association for Computing Machinery, New York, NY, USA, Article 17, 10 pages. doi:10.1145/2971485.2971528

[10] Hermann Ebbinghaus. 2013. Memory: a contribution to experimental psychology. *Annals of Neurosciences* 20, 4 (2013), 155–156. doi:10.5214/ans.0972.7531.200408

[11] Eva Eriksson, Gökçe Elif Baykal, Olof Torgersson, and Staffan Bjork. 2021. The CoCe Design Space: Exploring the Design Space for Co-Located Collaborative Games that Use Multi-Display Composition. In *Proceedings of the 2021 ACM Designing Interactive Systems Conference* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 718–733. doi:10.1145/3461778.3462023

[12] Maureen E. Estevez, Kristen A. Lindgren, and Peter R. Bergethon. 2010. A novel three-dimensional tool for teaching human neuroanatomy. *Anatomical sciences education* 3, 6 (2010), 309–317. doi:10.1002/ase.186

[13] Franz Faul, Edgar Erdfelder, Albert Georg Lang, and Axel Buchner. 2007. GPower 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191. doi:10.3758/BF03193146

[14] Paloma García-Robles, Irene Cortés-Pérez, Francisco Antonio Nieto-Escámez, Héctor García-López, Esteban Obrero-Gaitán, and María Catalina Osuna-Pérez. 2024. Immersive virtual reality and augmented reality in anatomy education: A systematic review and meta-analysis. *Anatomical Sciences Education* n/a, n/a (2024). doi:10.1002/ase.2397

[15] Fengming He, Xiyun Hu, Jingyu Shi, Xun Qian, Tianyi Wang, and Karthik Ramani. 2023. Ubi Edge: Authoring Edge-Based Opportunistic Tangible User Interfaces in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 461, 14 pages. doi:10.1145/3544548.3580704

[16] Kristina Höök, Peter Dalsgaard, Stuart Reeves, Jeffrey Bardzell, Jonas Löwgren, Erik Stolterman, and Yvonne Rogers. 2015. Knowledge Production in Interaction Design. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI EA '15)*. Association for Computing Machinery, New York, NY, USA, 2429–2432. doi:10.1145/2702613.2702653

[17] William Hue, Simon Jones, and Raj Sengupta. 2019. Exploring the Future Role of Self-Tracking Data in the Rheumatology Clinic. In *Healthcare of the Future (Studies in Health Technology and Informatics, Vol. 259)*, S. Bignens, T. Burkle, M. Lehmann, K. Denecke, E. Zetz, J. Holm, and M. Sariyar (Eds.). IOS Press, Netherlands, 33–38. doi:10.3233/978-1-61499-961-4-33

[18] IBM. 2024. *IBM SPSS Statistics*. SPSS. Retrieved February 16, 2024 from https://www.ibm.com/products/spss-statistics

[19] María Blanca Ibáñez, Ángela Di Serio, Diego Villarán, and Carlos Delgado Kloos. 2014. Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness. *Computers and education* 71 (2014), 1–13. doi:10.1016/j.compedu.2013.09.004

[20] Yvonne Jansen, Pierre Dragicevic, Petra Isenberg, Jason Alexander, Abhijit Karnik, Johan Kildal, Sriram Subramanian, and Kasper Hornbæk. 2015. Opportunities and Challenges for Data Physicalization. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3227–3236. doi:10.1145/2702123.2702180

[21] Jeffrey D Karpicke and Henry L. III Roediger. 2008. Critical Importance of Retrieval for Learning. *Science (American Association for the Advancement of Science)* 319, 5865 (2008), 966–968. doi:10.1126/science.1152408

[22] Pascal Knierim, Francisco Kiss, Maximilian Rauh, and Albrecht Schmidt. 2020. Tangibility is Overrated: Comparing Learning Experiences of Physical Setups and their Virtual Equivalent in Augmented Reality. In *Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia* (Essen, Germany) *(MUM '20)*. Association for Computing Machinery, New York, NY, USA, 299–305. doi:10.1145/3428361.3428379

[23] Timothy D. Lee. 1988. Chapter 7 Transfer-Appropriate Processing: A Framework for Conceptualizing Practice Effects in Motor Learning. In *Advances in Psychology*. Vol. 50. Elsevier Science & Technology, Amsterdam, 201–215. doi:10.1016/S0166-4115(08)62557-1

[24] Hui Liang, Chao Ge, Yusheng Sun, PU Li, Fei Liang, and CHANGHAI Wang. 2021. Prototype of porcelain safety display based on Augmented Reality Technology. In *Proceedings of the 2020 9th International Conference on Networks, Communication and Computing* (Tokyo, Japan) *(ICNCC '20)*. Association for Computing Machinery, New York, NY, USA, 57–64. doi:10.1145/3447654.3447663

[25] Jessica M. Logan, Andrew J. Thompson, and David W. Marshak. 2011. Testing to enhance retention in human anatomy. *Anatomical sciences education* 4, 5 (2011), 243–248. doi:10.1002/ase.250

[26] Lumivero. 2024. *NVivo*. Lumivero. Retrieved August 26, 2024 from https://lumivero.com/products/nvivo/

[27] Microsoft. 2024. *Direct manipulation with hands*. Microsoft. Retrieved March 21, 2024 from https://learn.microsoft.com/en-us/windows/mixed-reality/design/direct-manipulation

[28] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Hamburg</city>, <country>Germany</country>, </conf-loc>) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 459, 15 pages. doi:10.1145/3544548.3581449

[29] C. Donald Morris, John D. Bransford, and Jeffery J. Franks. 1977. Levels of processing versus transfer appropriate processing. *Journal of verbal learning and verbal behavior* 16, 5 (1977), 519–533. doi:10.1016/S0022-5371(77)80016-9

[30] Thomas Muender, Anke Verena Reinschluessel, Daniela Salzmann, Thomas Lück, Andrea Schenk, Dirk Weyhe, Tanja Döring, and Rainer Malaka. 2022. Evaluating Soft Organ-Shaped Tangibles for Medical Virtual Reality. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 237, 8 pages. doi:10.1145/3491101.3519715

[31] Tamotsu Murakami, Kazuhiko Hayashi, Kazuhiro Oikawa, and Naomasa Nakajima. 1995. DO-IT: Deformable Objects as Input Tools. In *Conference Companion on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '95)*. Association for Computing Machinery, New York, NY, USA, 87–88. doi:10.1145/223355.223442

[32] Magdalena Novak and Stephan Schwan. 2021. Does Touching Real Objects Affect Learning? *Educational Psychology Review* 33, 2 (2021), 637–665. doi:10.1007/s10648-020-09551-z

[33] Claire O'Malley and Danae Stanton Fraser. 2004. Literature Review in Learning with Tangible Technologies.

[34] Egidijus Pelanis, Rahul P Kumar, Davit L Aghayan, Rafael Palomar, Åsmund A Fretland, Henrik Brun, Ole Jakob Elle, and Bjørn Edwin. 2020. Use of mixed reality for improved spatial understanding of liver anatomy. *Minimally Invasive Therapy & Allied Technologies* 29, 3 (2020), 154–160.

[35] Daniel Preece, Sarah B. Williams, Richard Lam, and Renate Weller. 2013. "Let's Get Physical": Advantages of a physical model over 3D computer models and textbooks in learning imaging anatomy. *Anatomical Sciences Education* 6, 4 (2013), 216–224. doi:10.1002/ase.1345 arXiv:https://anatomypubs.onlinelibrary.wiley.com/doi/pdf/10.1002/ase.1345

[36] PTC. 2024. *Vuforia*. PTC. Retrieved February 7, 2024 from https://www.ptc.com/en/products/vuforia

[37] Anke Verena Reinschluessel, Thomas Muender, Tanja Döring, Verena Nicole Uslar, Thomas Lück, Dirk Weyhe, Andrea Schenk, and Rainer Malaka. 2021. A Study on the Size of Tangible Organ-shaped Controllers for Exploring Medical Data in VR. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 261, 7 pages. doi:10.1145/3411763.3451594

[38] Imi Ridley. 2024. *ImMed Illustration*. ImMed Illustration. Retrieved February 19, 2024 from imiridley.com

[39] Henry L. Roediger and Jeffrey D. Karpicke. 2006. Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological science* 17, 3 (2006), 249–255. doi:10.1111/j.1467-9280.2006.01693.x

[40] Jarek Rossignac. 2005. Shape complexity. *The Visual computer* 21, 12 (2005), 985–996.

[41] Kadek Ananta Satriadi, Jim Smiley, Barrett Ens, Maxime Cordeil, Tobias Czauderna, Benjamin Lee, Ying Yang, Tim Dwyer, and Bernhard Jenny. 2022. Tangible Globes for Data Visualisation in Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 505, 16 pages. doi:10.1145/3491102.3517715

[42] Orit Shaer, Eva Hornecker, et al. 2010. Tangible user interfaces: past, present, and future directions. *Foundations and Trends® in Human–Computer Interaction* 3, 1–2 (2010), 4–137.

[43] Peter Sommerauer and Oliver Müller. 2014. Augmented reality in informal learning environments: A field experiment in a mathematics exhibition. *Computers and education* 79 (2014), 59–68. doi:10.1016/j.compedu.2014.07.013

[44] Andrew T. Stull, Morgan J. Gainer, and Mary Hegarty. 2018. Learning by enacting: The role of embodiment in chemistry education. *Learning and instruction* 55

(2018), 80–92. doi:10.1016/j.learninstruc.2017.09.008

[45] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (2006), 237–246. doi:10.1177/1098214005283748 arXiv:https://doi.org/10.1177/1098214005283748

[46] Julie Urlings, Sümeyye Sezer, Mark ter Laan, Ronald Bartels, Thomas Maal, Jeroen Boogaarts, and Dylan Henssen. 2022. The role and effectiveness of augmented reality in patient education: A systematic review of the literature. *Patient Education and Counseling* 105, 7 (2022), 1917–1927. doi:10.1016/j.pec.2022.03.005

[47] Tim Vernon and Daniel Peckham. 2002. The benefits of 3D modelling and animation in medical teaching. *Journal of audiovisual media in medicine* 25, 4 (2002), 142–148. doi:10.1080/0140511021000051117

[48] Bruce Wainman, Liliana Wolak, Giancarlo Pukas, Eric Zheng, and Geoffrey R Norman. 2018. The superiority of three-dimensional physical models to two-dimensional computer presentations in anatomy learning. *Medical education* 52, 11 (2018), 1138–1146. doi:10.1111/medu.13683

[49] Tao Wang and Haibin Ling. 2018. Gracker: A Graph-Based Planar Object Tracker. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2018), 1494–1501. doi:10.1109/TPAMI.2017.2716350

[50] Maheshya Weerasinghe, Verena Biener, Jens Grubert, Aaron Quigley, Alice Toniolo, Klen Copic Pucihar, and Matjaz Kljun. 2022. VocabulARy: Learning Vocabulary in AR Supported By Keyword Visualisations. *IEEE transactions on visualization and computer graphics* 28, 11 (2022), 1–11. doi:10.1109/TVCG.2022.3203116

[51] Terri L. Woods, Sarah Reed, Sherry Hsi, John A. Woods, and Michael R. Woods. 2016. Pilot Study Using the Augmented Reality Sandbox to Teach Topographic Maps and Surficial Processes in Introductory Geology Labs. *Journal of Geoscience Education* 64, 3 (2016), 199–214. doi:10.5408/15-135.1 arXiv:https://doi.org/10.5408/15-135.1

[52] Kaissar Yammine and Claudio Violato. 2016. The effectiveness of physical models in teaching anatomy: a meta-analysis of comparative studies. 883–896 pages. doi:10.1007/s10459-015-9644-7