

Development of a Virtual Three-Dimensional Assessment Scenario for Anatomical Education

Katerina Bogomolova,^{1,2*} Amir H. Sam,³ Adam T. Misky,⁴ Chinmay M. Gupte,^{3,4} Paul H. Strutton,⁵ Thomas J. Hurkxkens,⁶ Beerend P. Hierck^{2,7,8}

¹Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands

²Center for Innovation of Medical Education, Leiden University Medical Center, Leiden, the Netherlands

³Medical Education Research Unit, Imperial College School of Medicine, Imperial College London, London, UK

⁴St Mary's Hospital, Imperial College Healthcare NHS Trust, London, UK

⁵Department of Surgery and Cancer, Imperial College London, London, UK

⁶Digital Learning Hub, Imperial College London, London, UK

⁷Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, the Netherlands

⁸Leiden Teachers' Academy, Leiden University, Leiden, the Netherlands

In anatomical education three-dimensional (3D) visualization technology allows for active and stereoscopic exploration of anatomy and can easily be adopted into medical curricula along with traditional 3D teaching methods. However, most often knowledge is still assessed with two-dimensional (2D) paper-and-pencil tests. To address the growing misalignment between learning and assessment, this viewpoint commentary highlights the development of a virtual 3D assessment scenario and perspectives from students and teachers on the use of this assessment tool: a 10-minute session of anatomical knowledge assessment with real-time interaction between assessor and examinee, both wearing a HoloLens and sharing the same stereoscopic 3D augmented reality model. Additionally, recommendations for future directions, including implementation, validation, logistic challenges, and cost-effectiveness, are provided. Continued collaboration between developers, researchers, teachers, and students is critical to advancing these processes. *Anat Sci Educ* 14: 385–393. © 2021 The Authors. *Anatomical Sciences Education* published by Wiley Periodicals LLC on behalf of American Association for Anatomy.

Key words: gross anatomy education; undergraduate education; stereoscopic three-dimensional visualization technology; formative assessment; summative assessment; augmented reality; virtual reality

*Correspondence to: Dr. Katerina Bogomolova, Department of Surgery, Leiden University Medical Center, PO Box 9600, Zone D6-P, 2300 RC Leiden, The Netherlands. E-mail: k.bogomolova@lumc.nl

Additional supporting information can be viewed in the online version of this article.

Drs. Katerina Bogomolova, Amir H. Sam, and Adam T. Misky contributed equally to this work and are considered joint first authors on this publication.

Received 13 March 2020; Revised 18 December 2020; Accepted 13 January 2021.

Published online 5 March 2021 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/ase.2055

© 2021 The Authors. *Anatomical Sciences Education* published by Wiley Periodicals LLC on behalf of American Association for Anatomy

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

INTRODUCTION

The use of three-dimensional visualization technology (3DVT), where interactive three-dimensional (3D) models are viewed on a two-dimensional (2D) screen, has been thoroughly explored in anatomical education and research (Yammine and Violato, 2015; Erolin et al., 2019). Due to its disadvantages for students with lower visual-spatial abilities, the focus of current research is gradually shifting toward 3DVT that is able to project the anatomical models in real 3D, that is, stereoscopically (Garg et al., 1999a; Huk, 2006; Luursema et al., 2008; Cui et al., 2017; Hackett and Proctor, 2018; Maresky et al., 2019; Bogomolova et al., 2020). Stereoscopic 3DVT allows learning and understanding of anatomical spatial relations, as traditional 3D teaching methods, such as plastinated specimens and cadaveric dissections, are becoming scarcer due to higher costs and decreased teaching time (Pryde and Black, 2005; Waterston and Stewart, 2005; Azer

and Eizenberg, 2007; Drake et al., 2009; Bergman et al., 2013). Consequently, many academic educational programs promote 3DVT for learning, to bridge the authenticity gap between using 2D content for education and achieving the competence in 3D spatial reasoning necessary for professional practice.

Stereoscopic 3DVT has been demonstrated to be effective in increasing the anatomical knowledge, and in stimulating learners' motivation and engagement (Luursema et al., 2008; Luursema et al., 2017; Moro et al., 2017; Ekstrand et al., 2018; Hackett and Proctor, 2018; Bogomolova et al., 2020; Wainman et al., 2020). Students with lower visual-spatial abilities appear to benefit most from learning within a stereoscopic 3D environment (Cui et al., 2017; Bogomolova et al., 2020). Such environments include virtual reality (VR) and stereoscopic augmented reality (AR). Stereovision in a VR environment is obtained with supportive devices such as Oculus Rift™ (Oculus VR, Menlo Park, CA) and HTC VIVE™ (High Tech Computer Corp., New Taipei City, Taiwan). A head-mounted display HoloLens™ (Microsoft Corp., Redmond, WA) is explicitly used in a stereoscopic AR environment. The particular strengths of stereoscopic AR lie in its ability to integrate the real environment with 3D virtual content that can be explored in "real 3D." This allows the user to manipulate and interact with virtual objects, while maintaining contact with their real environment, giving multiple users the ability to collaborate in real time, as well as to integrate the use of real specimen with (overlaid) virtual content (Pratt et al., 2018).

With emerging 3D learning initiatives, misalignment of learning with assessment is growing. While there is a notable shift in teaching methods from 2D to 3D methods, the assessment methods are shifting exactly the opposite way. Before the mid-2000s, practical examinations, such as potted specimens and prosections, were the most common assessment methods of anatomy in undergraduate medical education (Choudhury and Freemont, 2017). Today, medical students are more frequently assessed using written tests, such as multiple choice questions (MCQ), extended matching questions (EMQ), and single best answer questions (SBA), which are 2D in nature (Rowland et al., 2011; Choudhury and Freemont, 2017). The frequent use of written methods is mostly driven by their reduced need for temporal, financial, and human resources as compared to practical assessments (Rowland et al., 2011). Although, written tests, if well designed, are able to assess spatial understanding of anatomy, they lack spatial representation, and learning contextualization and engagement (McGrath et al., 2017). Not surprisingly, practical examinations are still identified as the most preferred method for assessment of anatomical knowledge among students, trainees, and specialists (Rowland et al., 2011).

The implementation of virtual 3D assessment methods has the potential to dissolve the growing misalignment between learning and assessment in undergraduate medical curricula. While there is plentiful available literature on the use of virtual 3D environments for teaching anatomy (Yamine and Violato, 2015; Erolin et al., 2019), its use in assessment is as yet unexplored. Virtual assessment scenarios currently reviewed in the literature focus primarily on clinical skills, rather than anatomy (Zackoff et al., 2019).

Here, the authors describe the development of a virtual 3D anatomy assessment and the perspectives from teachers and students on the use of this assessment tool in a medical curriculum. Design-based research (DBR) methodology was adopted for the purpose of this study, encompassing the phases of design, implementation, evaluation, and redesign of a virtual 3D assessment scenario, based on collaboration among students, teachers, learning experience designers, developers, and researchers in anatomical education at Imperial College

London, United Kingdom and Leiden University Medical Center, The Netherlands (Figure 1). Design-based research aims to improve educational practices through iterative analysis throughout the design, development, and implementation of the product (Wang and Hannafin, 2005; Dolmans and Tigelaar, 2012). Design-based research methodology has been widely used in higher education, especially for technological interventions (Anderson and Shattuck, 2012; Zheng, 2015; Rasouli et al., 2019; Zydney et al., 2020). It has been successfully described for design and development of mixed reality simulation for skills development in the paramedic field (Cowling and Birt, 2018). Since design and evaluation take place in a real educational context, notorious problems of integration from experimental context to real practice, that are common for technological interventions, do not occur. The DBR process took place in the period between May and July 2019. Thematic analysis was used to analyze the answers that students, trainees, and teachers have provided to the open-ended questions in the evaluation questionnaire (Braun and Clarke, 2006; Braun et al., 2019). Coding and theme development were performed in a deductive way using the existing concepts including the overall experience (usefulness, effectiveness, and enjoyment), technical innovation (3D visualization and dynamic exploration), and software difficulties (practical and technical issues). Appropriate ethical approval was granted at Imperial College London (registration no. MEEC1819-157).

VIRTUAL 3D ANATOMY ASSESSMENT

For the virtual 3D anatomy assessment scenario (see Figure 1), three distinct components were designed, implemented, and evaluated simultaneously: (1) the content of a 3D anatomy test, (2) the augmented reality application, and (3) the virtual 3D assessment scenario that adds the practical application in the assessment setting.

Stakeholder sessions were organized to design, evaluate, and redesign the 3D assessment scenario. Stakeholders included medical undergraduates, postgraduate trainees, anatomy teachers as content owners, learning and assessment designers, developers, and researchers.

The development of the anatomy test took place within two iterations of design and evaluation. In the first cycle, the anatomy test was designed by a team of four teachers and two researchers. Anatomy questions included low- and high-order thinking questions according to the Blooming Anatomy Tool (Bloom et al., 1956; Thompson and O'Loughlin, 2015). Additionally, questions were focused on subject matters that are difficult to assess on paper. Content validation was performed by five experts in the field of anatomy. Based on their feedback, the same team of teachers and researchers redesigned the questions during the second iterative cycle. Thereafter, the redesigned anatomy test was evaluated for clarity by four postgraduate trainees and two anatomy teachers. The final test consisted of 13 questions requiring identification of structures of the lower leg, determining their spatial relationships and functions, and indicating impaired functions in a clinical scenario (Supporting Information File 1). All questions aimed to assess spatial anatomical knowledge of the lower leg and functional aspects of the ankle joint (Supporting Information File 1).

An AR application for HoloLens®, Version 1, (Microsoft, Corp., Redmond, WA) was developed to integrate the anatomy test and anatomical 3D model into a virtual 3D assessment scenario (Figure 2). The design was based on the *DynamicAnatomy*® (Leiden University Medical Center, Leiden, The Netherlands)

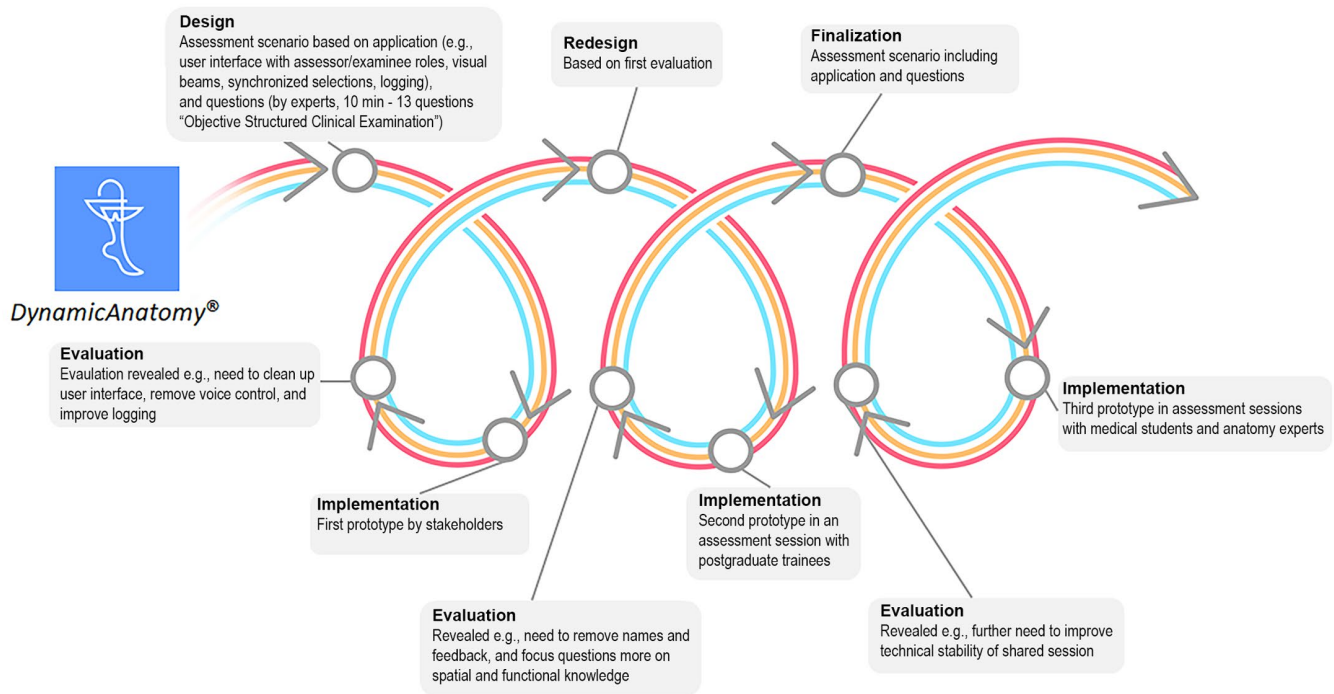


Figure 1.

Schematic representation of the design-based research approach. Multiple iterative cycles of design, implementation, evaluation, and redesign took place to finalize the virtual three-dimensional (3D) assessment. The three lines (red, yellow, and blue) represent the development trajectory of the anatomy test, augmented reality application, and 3D assessment scenario. All three components were designed, implemented, and evaluated simultaneously.

application, a fully interactive application focused on a stereoscopic 3D model of the lower leg that is presented as a 3D virtual object in a physical space (Bogomolova et al., 2020). Dynamic exploration allowed users to walk around the model and explore it from all possible angles. By sharing the same model and user interface among two or more users collaborative learning was stimulated. The interactive user control allowed virtual dissection of the leg by showing and hiding anatomical structures, categorized in skeletal, muscular, nervous, vascular and connective tissue systems, or structures individually. The four basic ankle movements were displayed as interactive 3D animations. Additionally, users were able to move the ankle joints manually. All structures and movements were selected and highlighted from a list by hand gestures or voice commands.

A team of students, researchers, and teachers evaluated the application through multiple iterative cycles in order to simplify and improve the use of application. Based on their feedback, redesign was mainly focused on improving utility, communication between examinee and assessor, and compliance to formal assessment rules and regulations. Improving utility focused on simplifying the menu options and eliminating visual and auditory feedback that were unnecessary and distracting in both the examinee's, and the assessor's user interface. Communication aspects focused on visualization of the examinee's and assessor's viewing and pointing directions, and synchronization of selected structures or movements among the devices. Compliance aspects focused on logging and recording of questions asked by the assessor and answers which were confirmed by the examinee in a logfile and on video.

The virtual 3D assessment scenario was a 10-minute session with real-time interaction between the assessor and examinee, both wearing a HoloLens and sharing the same stereoscopic

3D AR model (Figure 3). Next to the model, several tabs were projected including a list of ankle movements and confirmation button for the selected answer. A passive observer, also wearing a HoloLens, logged, and filmed the assessment for formal procedural purposes. The assessor called out the questions one by one allowing the student to spend as much time as wanted on each question, although within the confines of the maximum total time. To provide the answer, student selected the appropriate structure in the model or the type of movement from the list. To confirm the answer, students selected the confirmation button. Only after the answer was manually confirmed by the examinee, did the assessor proceed to the next question.

After first redesign, the virtual 3D assessment scenario was tested by six undergraduate medical students. In the final iterative cycle, the assessment scenario was evaluated by four students and six teachers. Feedback regarding personal experience and practical use was collected through a standardized self-reported questionnaire (Supporting Information File 2). Personal experience included items on enjoyment, satisfaction, preference, and perceived effectiveness of this 3D assessment. Practical use included items on usability of the application, quality of the stereoscopic 3D AR model, and experienced discomfort wearing the device. Open boxes were included for additional feedback on overall experience and/or suggestions for future implementations. The session took place in a 15-20 m² room where all participants could move freely, without any limitations. All participants were able to complete the anatomical test within 10 minutes. Some reported temporary loss of synchronization during the session. This problem was resolved immediately by the technician during the session. The provided feedback regarding personal experience and practical use during the evaluation sessions is presented in Tables 1 and 2.

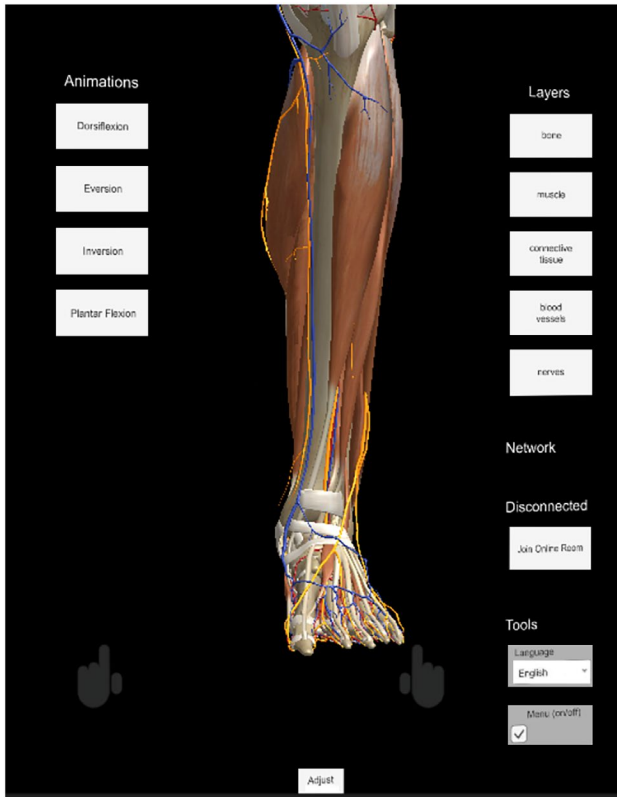


Figure 2.

A screenshot of augmented reality (AR) application from the user's point of view before the official start of the assessment.

DISCUSSION

Student's and Trainee's Perspectives

In general, the overall experience with the 3D virtual assessment among students and trainees was positive. Students and trainees reported a preference of the virtual 3D assessment over cadaveric/specimen assessment. Participants also highlighted this assessment's potential for greater standardization of the assessment material: *"I think it standardizes the assessment, but theoretically all cadavers should have the same components."* Teachers reported the benefits of virtual dissection to identify deep anatomical structures as the largest gain when compared to cadaveric examinations: *"I think the same aspects can be tested on a cadaver, but the HoloLens allows you to ask more questions a lot more easily. You can very quickly switch between anatomical aspects (bone/nerve/muscle) which can be different among multiple prosections."*

Several themes and subthemes were identified in thematic analysis of open-ended questions in the evaluation questionnaire (Table 3, Figure 4). In their feedback regarding the *Assessment Experience*, participants commented specifically on the usefulness of this assessment modality, its effectiveness in allowing the students to demonstrate knowledge, and how much they enjoyed the assessment as well as the interaction with the assessor. Regarding *Technical Innovation aspects of assessment*, participants particularly commented on its interactive 3D visualization technology, the ability to perform a virtual dissection within the software as well as the ease of exploring the 3D environment generated by the assessment. *Software Difficulties* included subthemes *Practical Issues* and *Technical Issues*. Some issues with the software used for the assessment were revealed during testing of the assessment. No major problems with the hardware were identified.

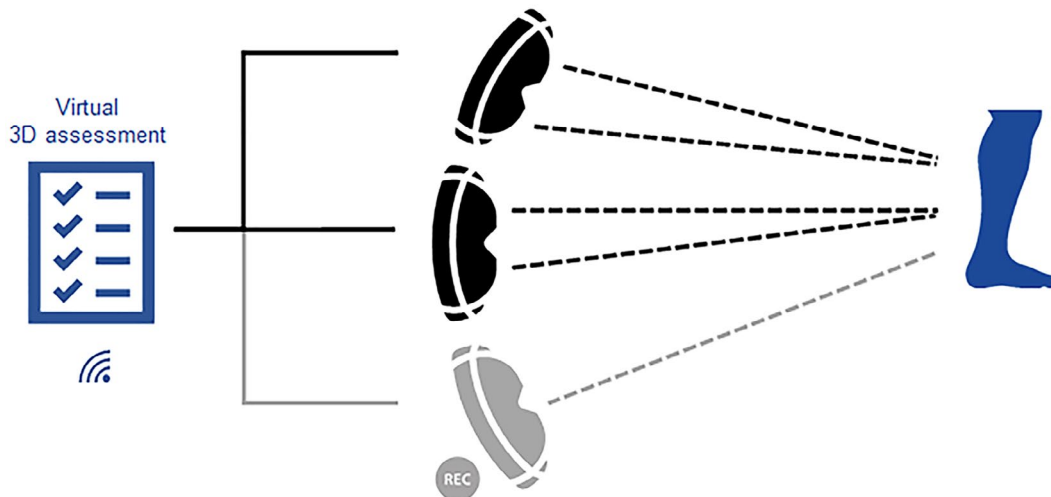


Figure 3.

Schematic representation of the assessment scenario. Three individuals wearing HoloLenses were participating in the assessment. The assessor and examinee (black) both shared a stereoscopic view (double dashed line) of the holographic three-dimensional (3D) model of the lower leg including user interface. A passive observer (gray), also wearing a HoloLens, shared the same view but monoscopically (single dashed line) and recorded the session on video. All devices were synchronized through a local Wi-Fi network, and all devices stored a logfile with timed registrations of operations and confirmed answers.

Table 1.

Personal Experience with the Virtual Three-Dimensional (3D) Assessment Scenario

Statement	Undergraduate students (n = 4) mean (\pm SD)	Postgraduate students (n = 6) mean (\pm SD)	Experts (n = 6) mean (\pm SD)
I enjoyed this assessment method	1.25 (\pm 0.50)	1.17 (\pm 0.41)	1.75 (\pm 0.96)
I enjoyed the interaction with the examiner	1.50 (\pm 0.58)	1.17 (\pm 0.41)	1.00 (\pm 0.00)
I could demonstrate my knowledge effectively	2.0 (\pm 0.82)	1.50 (\pm 0.55)	1.33 (\pm 0.52)
I prefer 3D assessment with HoloLens over paper-based assessment I have previously experienced	1.75 (\pm 0.96)	1.17 (\pm 0.41)	1.67 (\pm 0.82)
Certain anatomical aspects can be assessed in 3D, but not in 2D	2.75 (\pm 1.71)	1.17 (\pm 0.41)	1.5 (\pm 0.55)
I prefer 3D assessment with HoloLens over 3D assessment on a cadaver	2.00 (\pm 0.00)	1.67 (\pm 0.52)	2.33 (\pm 0.82)
Certain anatomical aspects can be assessed in 3D with HoloLens, but not on a cadaver	2.0 (\pm 1.41)	2.17 (\pm 0.98)	1.5 (\pm 0.55)
I prefer 3D assessment with HoloLens over 3D assessment on a prosection	2.25 (\pm 0.96)	1.50 (\pm 0.55)	2.33 (\pm 0.52)
Certain anatomical aspects can be assessed in 3D with HoloLens, but not on a prosection	3.00 (\pm 1.15)	2.50 (\pm 1.38)	1.6 (\pm 0.55)
I feel more confident about my anatomical competences after a HoloLens examination (compared to paper-based examination)	1.75 (\pm 0.50)	1.33 (\pm 0.52)	1.75 (\pm 0.5)
I would study differently knowing that I will be assessed with the HoloLens	1.50 (\pm 1.00)	2.17 (\pm 1.33)	2.5 (\pm 1.38)

Response options on a five-point Likert scale ranged from 1 = strongly agree to 5 = strongly disagree. Average scores are expressed in means (\pm SD).

Table 2.

Feedback Regarding Practical Use of the Application and Device

Statement	Undergraduate students (n = 4) mean (\pm SD)	Postgraduate students (n = 6) mean (\pm SD)	Experts (n = 6) mean (\pm SD)
The quality of the holographic model was adequate	1.75 (\pm 0.96)	1.33 (\pm 0.52)	1.67 (\pm 0.52)
The application was easy to use	2.50 (\pm 1.73)	2.17 (\pm 1.33)	2.0 (\pm 0.63)
I experienced discomfort wearing the HoloLens	3.0 (\pm 1.83)	3.83 (\pm 1.47)	4.0 (\pm 1.26)
The HoloLens hindered me in giving the answers to the questions	4.0 (\pm 1.41)	4.4 (\pm 1.34)	4.33 (\pm 1.21)

Response options on a five-point Likert scale ranged from 1 = strongly agree to 5 = strongly disagree. Average scores are expressed in means (\pm SD).

Potential Benefits of 3D Assessment

The benefits of virtual 3D assessments are manifold. First, like traditional practical examinations, it allows for testing 3D spatial understanding, anatomical variations, and understanding of relations of structures to each other (Smith and

McManus, 2015). Second, the assessment in a virtual 3D environment can allow students, especially those with lower visual-spatial abilities, to demonstrate their knowledge more effectively. In general, students with lower visual spatial abilities experience difficulties with translating 2D images into 3D mental representations and vice versa (Garg et al., 1999a, b,

Table 3.

Summary of Themes and Subthemes form Thematic Analysis of Participants' Feedback with Quotes Evidencing and Highlighting the Subthemes

Themes/Subthemes	Students quotes
Positive Assessment Experience	
Usefulness	<ul style="list-style-type: none"> • “A potentially very useful way of understanding” • “Can see its real practical use and a good ease of learning” • “Extremely useful! Can’t wait till it becomes more commercial” • “Good for formative assessment with library based HoloLens” • “Has potential, especially in small cohorts. Logistics need to be addressed”
Effectiveness	<ul style="list-style-type: none"> • “Good for allowing a variety of questions on a single specimen” • “Easy to test multiple aspects of anatomy quickly with easy switching between questions” • “Allows mixture of functional and gross anatomical questions”
Enjoyment	<ul style="list-style-type: none"> • “Interactive and enjoying” • “Fun. Didn’t smell like cadaver” • “An interesting and potentially very rewarding experience” • “I am very pleased with this kind of assessment in the future”
Technical Innovation	
Novelty	<ul style="list-style-type: none"> • “Novel and engaging” • “Really cool way of assessing” • “I liked the novelty of it” • “Very innovative and useful way to view anatomy”
Three-dimensional (3D) Visualization	<ul style="list-style-type: none"> • “I liked the 3 dimensions offered by the glasses and app” • “I like the interactivity aspect and the fact that the anatomy was easy to visualize, something which is lacking through 2D textbook teaching and revision at the moment” • “For me the depth perspective is very important. This is not possible from books”
Virtual Dissection	<ul style="list-style-type: none"> • “Easy separation of structures and reset in space for learning is possible in HoloLens but not on a cadaver” • “Virtual dissection is possible. Also the multiple viewing angles are easier compared to Cadaver” • “Deep structures in particular are more accessible” • “I liked being able to add/remove layers and highlight whole muscles”
Dynamic Exploration	<ul style="list-style-type: none"> • “Liked being able to walk around object freely” • “How you could move around and see it in different angles was great” • “Overall it was easy to use, being able to walk around the limb was useful as it re-created the experience of a real limb exam”
Software Difficulties	
Practical Issues	<ul style="list-style-type: none"> • “Clean interface, easy to strip items and look from multiple angles” • “Very clear and easy to me. Sometimes frustrating (couldn’t click on nerves)” • “It was difficult to highlight smaller landmarks (nerves/blood vessels) compare to larger ones (muscles/bones)”
Technical Issues	<ul style="list-style-type: none"> • “Overall positive. More experience with software would improve process” • “Crashed during assessment - might affect confidence in exam” • “I think in concept HoloLens would be useful in ensuring sound anatomical, knowledge, however, at the moment the technical difficulties and the anxieties these would induce in an exam are not favorable”

2002; Huk, 2006). Since these cognitive steps of mental translation are not required in a virtual 3D environment, students will be able to allocate more cognitive resources to visualization and comprehension of spatial anatomy (Hegarty and Sims, 1994; Bulthoff et al., 1995; Garg et al., 2001; Huk,

2006; Mayer, 2014; Bogomolova et al., 2020). Third, it can stimulate deep learning, since assessment methods are able to influence and drive the learning approach students take (Newble and Gordon, 1985; Reid et al., 2007). Fourth, it creates a so-called collaborative virtual environment (Lacoché

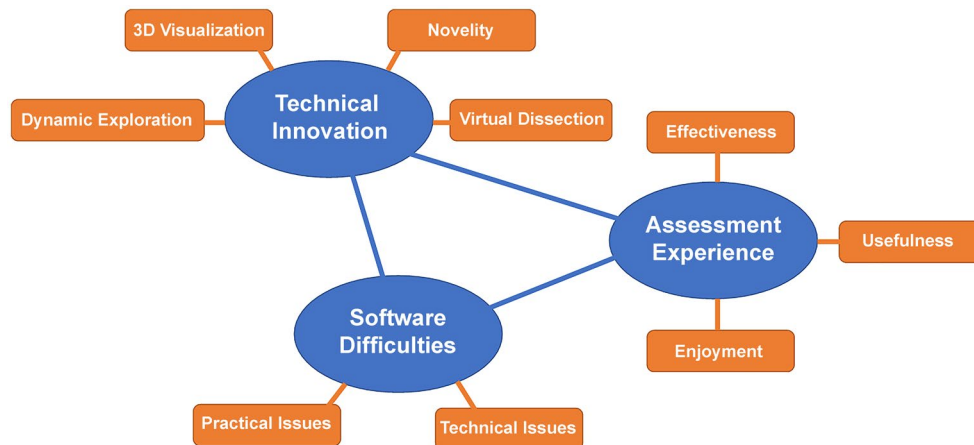


Figure 4.

Schematic map of themes and subthemes from thematic analysis of participants' feedback on the novel three-dimensional (3D) anatomy assessment.

et al., 2017). Stereoscopic 3D AR environment in particular, allows interaction between teacher and student who share and manipulate the same virtual object, while maintaining contact with their real environment. Fifth, it guarantees fairness in the assessment when compared to practical examinations. While potted specimens and prosections include anatomical variations, the standardized 3D assessment will remain identical for each student. Lastly, it can further promote the implementation of 3D education and allow medical undergraduates and junior doctors to overcome difficulties in translating acquired anatomical knowledge from their studies into clinical practice (McKeown et al., 2003; Prince et al., 2005; Spielmann and Oliver, 2005; Waterston and Stewart, 2005; Bergman et al., 2008).

The above benefits were clearly noted by students and teachers during the final evaluation sessions. Students perceived 3D assessment as fair because of content that is better standardized compared to cadaveric material. Additionally, they reported feeling more confident about their anatomical knowledge after the assessment and better prepared for a summative assessment. Teachers valued the flexibility of the scenario to assess knowledge and competences on all levels of (3D) complexity and specificity (Bloom et al., 1956). Students also reported that virtual 3D assessment could probably influence their approach to learning. While written examinations often stimulate surface learning, 3D assessment has the potential to stimulate deep learning among students (Newble and Gordon, 1985; Reid et al., 2007).

The practical advantage of an AR application-based assessment lies in its compliance with rules and regulations of formal assessments. It enables live automated registration of questions and answers, and provides data for psychometric evaluation to ensure validity of the examination (Norcini et al., 2018). This is important and necessary in times of decreasing temporal and human resources. Another benefit is the ability to incorporate real time, embedded feedback that can be used in formative assessments. Real-time feedback, as the central component of effective formative assessment, provides information about the existing gap between the actual and desired levels of knowledge (Al-Kadri et al., 2012; Mogali et al., 2020). This in turn, is likely to contribute to students' deep approach to learning (Rushton,

2009). Concurrently, students will get familiar with the assessment method and will be prepared for the summative part of it.

Future Directions

While virtual 3D assessment is clearly promising, there are still important issues that need to be addressed for successful implementation in medical curricula.

Within the context of proper assessment, care should be taken regarding the validity of the test in its unique assessment environment. In the current project, the anatomy test was designed and tested for research purposes only. However, as part of final implementation, a rigorous validation of the virtual 3D assessment is required to demonstrate the appropriateness of the interpretations, uses, and decisions based on assessment results (Cook and Hatala, 2016). According to the validity framework for simulation-based assessments, validation should be preferably built upon five sources of evidence: content, internal structure, relationship with other variables, response process, and consequences (Cook and Hatala, 2016). While assessment of content and internal structure can be straightforward, assessment of the relationships with other variables can remain challenging. These can include written tests of practical examinations that measure the same construct and are used as the standard assessment method (Parsons et al., 2008; Parsons and Rizzo, 2008; Waldman et al., 2008). However, as has been reported by various validation studies, the correlation between new assessment and standard methods does not always appear to be strong, even if the new method produces valid results (Waldman et al., 2008). This shows how challenging it can be to validate 3D assessment using other methods of assessment. Another way to assess the relationship with other variables, is to do expert-novice comparisons (Cook and Hatala, 2016). The consequences of 3D assessment can be evaluated in terms of student's motivation, changes in their approach to learning and clinical performance.

One of the logistic challenges include preservation of a flawless operation of the application and device. This is essential for a continuous flow of the assessment and providing confidence for both the assessor and the examinee. In the above case a small

number of technical issues stemmed from the connectivity stability of devices in the assessment session. This problem would need to be addressed attentively by design and development teams in all future AR assessment modalities. Another important issue lies in terms of scale, effort, and availability of both headsets/hardware and manpower. In our case, at least two devices, one for the teacher and one for the student, are needed for one assessment to take place. Since the amount of available devices is usually limited due to high costs, 3D assessment can be first implemented as part of formative assessment. This will allow 3D self-assessment within small group sessions with immediate feedback. Another way to remain within scale restrictions, is to balance the benefits of 3D virtual and 2D paper-and-pencil assessment. Blending the two methods could combine their strengths. Alternatives for AR/HoloLens could be found in for example, phone- or tablet-based ARKit (Apple Inc., Cupertino, CA) or ARCore Technology (Google LLC., Mountain View, CA), which could also help with scalability issues. Furthermore, scalability issues are also present in terms of manpower required to run such assessments. Participants in the above presented case particularly enjoyed the real-time interaction with the assessor in a small-scale study such as ours, one-on-one interaction is feasible, however when scaled to the size of entire undergraduate or postgraduate year groups, having a one-on-one assessment session with a single tutor quickly becomes unfeasible. In order to overcome this problem while maintaining the personal interaction our participants particularly enjoyed, future researchers could consider streaming a single instructor/assessor's presence to multiple students at the same time, or incorporate adaptive learning through the use of algorithms that generate responses based on the learners' input (Zackoff et al., 2019). Clearly further work is required to overcome this barrier.

Finally, cost-effectiveness should be considered in development and final implementation of virtual 3D assessment platform in curriculum. Startup resources must be available for continuous development and maintenance of content, and data management. Additionally, specific training for faculty will be needed to enable them to effectively operate and implement virtual 3D assessment platform. For such training programs significant amounts of time and resources can be required.

Limitations of the Study

Several limitations of this study were identified by the authors. First, the development of the anatomy test and AR application was based on the anatomy of the lower leg. Anatomical regions with a much greater complexity, such as brain, can cause new technical issues and different personal experiences. Therefore, development processes can vary in length and complexity, and can lead to different outcomes for particular anatomical regions. Second, the 3D assessment scenario was developed and tested in two institutions that share similar personal values and experiences. However, in parts of the world where traditions, personal values and experience are different, the proposed developmental approach and recommendations for the use of virtual 3D assessment in medical education may not suit their needs.

CONCLUSIONS

Based on our experience with the above study, as well as our review of the available literature, authors believe virtual 3D assessment can address several challenging aspects of the growing misalignment between learning, assessment, and clinical

practice. The development of a virtual 3D assessment scenario was successfully demonstrated using the theoretical framework of design-based research. For further implementation of virtual 3D assessment platform in medical education, rigorous validation, logistic challenges, and cost-effectiveness must be addressed carefully. Continued collaboration between developers, researchers, teachers, and students is critical to advancing these processes.

ACKNOWLEDGMENTS

The authors thank Prof. Marco de Ruyter from the Department of Anatomy and Embryology from Leiden University Medical Center for content validation, Roland Smeenk (Inspark), and Robin Kruyt (LUMC, Leiden, Netherlands) for application development and testing. *DynamicAnatomy*[®], on which this assessment application is based has been developed by Leiden University Medical Center and Leiden University Centre for Innovation.

NOTES ON CONTRIBUTORS

KATERINA BOGOMOLOVA, M.D., is a graduate (Ph.D.) student in the Department of Surgery and Center for Innovation in Medical Education at Leiden University Medical Center, in Leiden, The Netherlands. She is investigating the role of three-dimensional visualization technologies in anatomical and surgical education in relation to learners' spatial abilities.

AMIR H. SAM, M.B.B.S., Ph.D., is Head of Imperial College School of Medicine, in London, UK and a consultant physician and endocrinologist at Charing Cross and Hammersmith Hospitals in London, UK. His medical education research focuses on novel approaches to assessment of applied knowledge and clinical and professional skills.

ADAM T. MISKY, M.B.B.S., M.R.C.S., is a surgical trainee at the Royal Free London NHS Foundation Trust in London, UK. He is an active and keen medical educator and has special interest in novel undergraduate assessment methods and students' preconceptions on medical training.

CHINMAY M. GUPTE, M.A., Ph.D., F.R.C.S., B.M.B.Ch., is a consultant orthopedic surgeon and clinical reader (associate professor) in musculoskeletal science at Imperial College NHS Trust and Imperial College in London, UK. He leads the orthopedic virtual reality and simulation center at Imperial College.

PAUL H. STRUTTON, Ph.D., is a senior lecturer in neurophysiology in the Department of Surgery and Cancer at Imperial College London in London, UK. He teaches neuroscience and anatomy on medical and biomedical science degrees and supervises undergraduates and postgraduates undertaking research in his laboratory. His research focusses on the neural control of movement in health and disease.

THOMAS J. HURKXKENS, M.Sc., is lead of new media at the Digital Learning Hub of Imperial College London, in London, UK. His career focus is on innovation management, innovative digital learning tools, and the use of virtual and augmented reality for higher education.

BEEREND P. HIERCK, Ph.D., is an assistant professor of anatomy in the Department of Anatomy and Embryology and researcher at the Center for Innovation in Medical Education at Leiden University Medical Center in Leiden, The Netherlands. He teaches anatomy, developmental biology, and histology to (bio)medical students and his educational research focuses

on 3D learning in the (bio)medical curriculum, with a special interest in the use of extended reality.

LITERATURE CITED

- Al-Kadri HM, Al-moamary MS, Roberts C, van der Vleuten CP. 2012. Exploring assessment factors contributing to students' study strategies: Literature review. *Med Teach* 34:42–50.
- Anderson T, Shattuck J. 2012. Design-based research: A decade of progress in education research? *Educ Res* 41:16–25.
- Azer SA, Eizenberg N. 2007. Do we need dissection in an integrated problem-based learning medical course? Perceptions of first- and second-year students. *Surg Radiol Anat* 29:173–180.
- Bergman EM, de Bruin AB, Herrler A, Verhrijen IW, Scherpbier AJ, van der Vleuten CP. 2013. Students' perceptions of anatomy across the undergraduate problem-based learning medical curriculum: A phenomenographical study. *BMC Med Educ* 13:152–162.
- Bergman EM, Prince KJ, Drukker J, van der Vleuten CP, Scherpbier AJ. 2008. How much anatomy is enough? *Anat Sci Educ* 1:184–188.
- Bloom BS, Engelhart MD, Furst EJ, Hill WH, Krathwohl DR. 1956. *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook I: Cognitive Domain*. 1st Ed. New York, NY: David McKay Company. 207 p.
- Bogomolova K, van der Ham IJ, Dankbaar ME, van den Broek WW, Hovius SE, van der Hage JA, Hierck BP. 2020. The effect of stereoscopic augmented reality visualization on learning anatomy and the modifying effect of visual-spatial abilities: A double-center randomized controlled trial. *Anat Sci Educ* 13:558–567.
- Braun V, Clarke V. 2006. Using thematic analysis in psychology. *Qual Res Psychol* 3:77–101.
- Braun V, Clarke V, Hayfield N, Terry G. 2019. Thematic analysis. In: Liampittong P (Editor). *Handbook of Research Methods in Health Social Sciences*. 1st Ed. Singapore, Singapore: Springer Nature Singapore Pte Ltd. p 843–860.
- Bulthoff HH, Edelman SY, Tarr MJ. 1995. How are three-dimensional objects represented in the brain? *Cereb Cortex* 5:247–260.
- Choudhury B, Freemont A. 2017. Assessment of anatomical knowledge: Approaches taken by higher education institutions. *Clin Anat* 30:290–299.
- Cook DA, Hatala R. 2016. Validation of educational assessments: A primer for simulation and beyond. *Adv Simul (Lond)* 1:31.
- Cowling M, Birt J. 2018. Pedagogy before technology: A design-based research approach to enhancing skills development in paramedic science using mixed reality. *Information* 9:29.
- Cui D, Wilson TD, Rockhold RW, Lehman MN, Lynch JC. 2017. Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anat Sci Educ* 10:34–45.
- Dolmans DH, Tigelaar D. 2012. Building bridges between theory and practice in medical education using a design-based research approach: AMEE Guide No. 60. *Med Teach* 34:1–10.
- Drake RL, McBride JM, Lachman N, Pawlina W. 2009. Medical education in the anatomical sciences: The winds of change continue to blow. *Anat Sci Educ* 2:253–259.
- Ekstrand C, Jamal A, Nguyen R, Kudryk A, Mann J, Mendez I. 2018. Immersive and interactive virtual reality to improve learning and retention of neuroanatomy in medical students: A randomized controlled study. *CMAJ Open* 6: E103–E109.
- Erolin C, Reid L, McDougall S. 2019. Using virtual reality to complement and enhance anatomy education. *J Vis Commun Med* 42:93–101.
- Garg A, Norman GR, Spero L, Maheshwari P. 1999a. Do virtual computer models hinder anatomy learning? *Acad Med* 74:S87–S89.
- Garg A, Norman G, Spero L, Taylor I. 1999b. Learning anatomy: Do new computer models improve spatial understanding? *Med Teach* 21:519–522.
- Garg AX, Norman G, Sperotable L. 2001. How medical students learn spatial anatomy. *Lancet* 357:363–364.
- Garg AX, Norman GR, Eva KW, Spero L, Sharan S. 2002. Is there any real virtue of virtual reality? The minor role of multiple orientations in learning anatomy from computers. *Acad Med* 77:S97–S99.
- Hackett M, Proctor M. 2018. The effect of autostereoscopic holograms on anatomical knowledge: A randomized trial. *Med Educ* 52:1147–1155.
- Hegarty M, Sims VK. 1994. Individual differences in use of diagrams as external memory in mechanical reasoning. *Mem Cognit* 22:411–430.
- Huk T. 2006. Who benefits from learning with 3D models? The case of spatial ability. *J Comput Assist Learn* 22:392–404.
- Lacoe J, Pallamin N, Boggini T, Royan J. 2017. Collaborators awareness for user cohabitation in co-located collaborative virtual environments. In: *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*; Göteborg, Sweden, 2017 November 8–10. Article 15. Association for Computing Machinery, New York, NY.
- Luursema JM, Verwey WB, Kommers PA, Annema JH. 2008. The role of stereopsis in virtual anatomical learning. *Interact Comput* 20:455–460.
- Luursema JM, Vorstenbosch M, Kooloos J. 2017. Stereopsis, visuospatial ability, and virtual reality in anatomy learning. *Anat Res Int* 2017:1493135.
- Maresky HS, Oikonomou A, Ali I, Ditzkofsky N, Pakkal M, Ballyk B. 2019. Virtual reality and cardiac anatomy: Exploring immersive three-dimensional cardiac imaging, a pilot study in undergraduate medical anatomy education. *Clin Anat* 32:238–243.
- Mayer RE. 2014. Cognitive theory of multimedia learning. In: Mayer RE (Editor). *Multimedia Learning*. 2nd Ed. Santa Barbara, CA: Cambridge University Press. p 43–47.
- McGrath JL, Taekman JM, Dev P, Danforth DR, Mohan D, Kman N, Crichlow A, Bond WF. 2017. Using virtual reality simulation environments to assess competence for emergency medicine learners. *Acad Emerg Med* 25:186–195.
- McKeown PP, Heylings DJ, Stevenson M, McKelvey KJ, Nixon JR, McCluskey DR. 2003. The impact of curricular change on medical students' knowledge of anatomy. *Med Educ* 37:954–961.
- Mogali SR, Rotgans JI, Rosby L, Ferenczi MA, Low Beer N. 2020. Summative and formative style anatomy practical examinations: Do they have impact on students' performance and drive for learning? *Anat Sci Educ* 13:581–590.
- Moro C, Štromberga Z, Raikos A, Stirling A. 2017. The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anat Sci Educ* 10:549–559.
- Newble DI, Gordon MI. 1985. The learning style of medical students. *Med Educ* 19:3–8.
- Norcini J, Anderson MB, Bollela V, Bollela V, Burch V, Costa MJ, Duvivier R, Hays R, Palacios Mackay MF, Roberts T, Swanson D. 2018. Consensus framework for good assessment. *Med Teach* 40:1102–1109.
- Parsons TD, Rizzo AA. 2008. Initial validation of a virtual environment for assessment of memory functioning: Virtual reality cognitive performance assessment test. *CyberPsychol Behav* 11:17–25.
- Parsons TD, Silva TM, Pair J, Rizzo AA. 2008. Virtual environment for assessment of neurocognitive functioning: Virtual reality cognitive performance assessment test. In: Westwood JD, Haluck RS, Hoffman HM, Mogel GT, Phillips R, Robb RA, Vosburgh KG (Editors). *Medicine Meets Virtual Reality 16: Parallel, Combinatorial, Convergent: NextMed by Design*. 1st Ed. Amsterdam, The Netherlands: IOS Press BV. p 351–356.
- Pratt P, Ives M, Lawton G, Simmons J, Radev N, Spyropoulou L, Amiras D. 2018. Through the HoloLens looking glass: Augmented reality for extremity reconstruction surgery using 3D vascular models with perforating vessels. *Eur Radiol Exp* 2:2.
- Prince KJ, Scherpbier AJ, Van Mameren H, Drukker J, van der Vleuten CP. 2005. Do students have sufficient knowledge of clinical anatomy? *Med Educ* 39:326–332.
- Pryde FR, Black SM. 2005. Anatomy in Scotland: 20 years of change. *Scott Med J* 50:96–98.
- Rasouli B, Aliabadi K, Aradkani SP, Ahmady S, Asgari M. 2019. Determining components of medical instructional design based on virtual reality by research synthesis. *J Med Educ Dev* 14:232–244.
- Reid WA, Duvall E, Evans P. 2007. Relationship between assessment results and approaches to learning and studying in year two medical students. *Med Educ* 41:754–762.
- Rowland S, Ahmed K, Davies DC, Ashrafian H, Patel V, Darzi A, Paraskeva PA, Athanasios T. 2011. Assessment of anatomical knowledge for clinical practice: Perceptions of clinicians and students. *Surg Radiol Anat* 33: 263–269.
- Rushon A. 2009. Formative assessment: A key to deep learning? *Med Teach* 27:509–513.
- Smith C, McManus B. 2015. The integrated anatomy practical paper: A robust assessment method for anatomy education today. *Anat Sci Educ* 8:63–73.
- Spielmann PM, Oliver CW. 2005. The carpal bones: A basic test of medical students and junior doctors' knowledge of anatomy. *Surgeon* 3:257–259.
- Thompson AR, O'Loughlin VD. 2015. The Blooming Anatomy Tool (BAT): A discipline-specific rubric for utilizing Bloom's taxonomy in the design and evaluation of assessments in the anatomical sciences. *Anat Sci Educ* 8:493–501.
- Wainman B, Pukas G, Wolak L, Mohanraj S, Norman GR. 2020. The critical role of stereopsis in virtual and mixed reality learning environments. *Anat Sci Educ* 13:398–405.
- Waldman UM, Gulich MS, Zeitler HP. 2008. Virtual patients for assessing medical students: Important aspects when considering the introduction of a new assessment format. *Med Teach* 30:17–24.
- Wang F, Hannafin MJ. 2005. Design-based research and technology-enhanced learning environments. *Educ Technol Res Dev* 53:5–23.
- Waterston SW, Stewart IJ. 2005. Survey of clinicians' attitudes to the anatomical teaching and knowledge of medical students. *Clin Anat* 18:380384.
- Yamine K, Violato C. 2015. A meta-analysis of the educational effectiveness of three-dimensional visualization technologies in teaching anatomy. *Anat Sci Educ* 8:525–538.
- Zackoff MW, Real FJ, Cruse B, Davis D, Klein M. 2019. Medical student perspectives on the use of immersive virtual reality for clinical assessment training. *Acad Pediatr* 19:849–851.
- Zheng L. 2015. A systematic literature review of design-based research from 2004 to 2013. *J Comput Educ* 2:399–420.
- Zydney JM, Warner Z, Angelone L. 2020. Learning through experience: Using design based research to redesign protocols for blended synchronous learning environments. *Comput Educ* 143:103678.