The University of Sheffield

Deconstructed Learning in Pre-Clinical Dental Education Using Virtual Reality Simulation

Ashley Charles Towers

Supervisory Team:

Prof. Nicolas Martin
Dr. Steve Maddock
Prof. Christopher Stokes

A report submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

in the

School of Clinical Dentistry

2025

Abstract

Aim

The aim of this project is to assess if Virtual Reality, as an operative dental education tool, results in superior learning (as measured via a transfer task) when it is employed to provide deconstructed part-task exercises instead of facilitating whole-task repetitive practice.

- Literature Review: Explore contemporary uses and applications of VR in operative dental education (Scoping Review).
- Design of a Task Analysis approach through which operative tasks can be deconstructed into their sub-component tasks and reveal their underlying cognitive structure.
- Application of the Task Analysis approach to deconstruct an operative procedure (caries removal and cavity preparation).
- Creation of task-specific VR simulator exercises from the identified sub-components.
- Randomised control trial to assess the interventions: Deconstructed part-task versus whole-task training using VR simulation.
- Evaluation of results using: 1) Quantitative measures of performance, retention and transfer to establish superiority. 2) Qualitative evaluation of participant opinion to explore themes relating to perceptions and acceptability of the experimental approach.
- Presentation of a recommended approach for Task Analysis of operative tasks to support learning using deconstructed VR simulator exercises.

Results

- Motor skills research and student opinion do not support the premise that VR simulation should replace traditional phantom head simulation.
- The theory of motor skills specificity should be integrated into teaching to optimise transfer to target contexts.
- Task Analysis can be applied to produce a consensus, revealing divergence in operative approaches and promoting discussions leading to an agreed institutional approach.

- Task Analysis is a powerful tool for uncovering the underlying cognitive basis of tasks which can be used to develop deconstructed exercises or other new learning materials.
- Deconstructed exercises manage the complexity of learning operative tasks and facilitate the optimisation of transfer to other contexts. Such exercises can be effective and are welcomed by learners.
- Deconstructed learning may result in superior retention of declarative facts and improved judgement when compared to whole-task teaching approaches.

Acknowledgements

Completing a PhD is an ambition that I have held for a very long time, and I am incredibly grateful for the opportunity to have done so. This could not have been done without the support and encouragement of so many people along the way.

I would like to express my deepest thanks to my supervisory team: Professor Nicolas Martin, Dr Steve Maddock, and Professor Christopher Stokes. Your suggestions, guidance, and robust discussions since the beginning of this journey are massively appreciated and have very much helped to shape this work.

I would also like to particularly thank Arnaud Cosson and the team at HRV Simulation for the generous support which enabled this research project and for the implementation of the new software features used in this work.

I am also grateful to my colleagues and friends at the School of Clinical Dentistry, who I will not name individually for fear of missing someone, but you know who you are, and I thank you for checking in on my progress, being a sounding board for ideas, or just keeping my spirits up throughout the many frustrations!

Lastly, words cannot express my gratitude for the support of my family, especially my parents, Jaqueline and Stuart, my wife, Laura, and our daughters, Tabitha and Matilda. Your unwavering love, support, and belief in me has kept me going throughout this long journey.

Finally, I would like to dedicate this work to the memory of my late sister, Felicity, without whom I would have never been in the position to carry it out. I will never hear you say it, but I will forever be your Herr Doktor Bruder.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Thesis Structure

Simulation, in one form or another, has been part of dental training for over 1500 years but since the late 1800s dental students have primarily trained using Phantom Head based simulation (Owen, 2016). Phantom Heads provide a mannequin into which artificial (often made of plastic) or real extracted teeth can be placed where the learner can practice their technique using real dental tools in a relatively safe environment without putting a patient at risk (Fugill, 2013). The adoption of Virtual Reality (VR) into this context was driven by the limitations of finding appropriate extracted teeth on which to practice a specific procedure, the limitations of plastic teeth to replicate realistic experiences, lack of tutor time to provide feedback and the subjectivity of tutor assessment (Xia et al., 2013).

If a perfect VR simulation of dentistry were possible, traditional training methods could be simply reapplied in the new context taking inspiration from current approaches in the clinical skills lab and those used with real patients. However, it is not possible to create a perfect simulation so compromises have to be made in both hardware and software. Unlike commercial flight simulators where a fully functional and accurate cockpit is provided, current dental simulator designers have settled on a core set of human interface devices that provide an approximate representation of the sensations encountered operatively: namely stereoscopic displays, haptic force feedback, foot pedals and simulated dental handpieces and hand tools (Wang et al., 2016a). These devices combine to provide only a subset of the real environment and exclude sensations such as the spray from the water cooling, the heat of the bur, or even the smells created as tissue is abrasively removed. Even the included aspects of the experience are an approximation; the tooth cutting process is frequently modelled as direct subtraction of tooth surface voxels where they intersect with a 3d model of the dental bur (e.g. Wu et al. (2010)) rather than modelling the complex abrasion, sheer and plastic deformations that occur when a real bur is operating upon dental tissue. Consequently, the subtle haptic feedback delivered through a real handpiece is often simplified to a ‘handpiece on' vibration and a relatively small number of levels of physical resistance representing the different dental tissues.

This simplification of the simulation has, arguably, extended to the scoring and assessment mechanisms provided by the simulator. The most common scoring mechanism used by dental simulators informs the user what percentage of a target shape they have removed and how much iatrogenic damage was caused to the area outside of the target in the process. Where an exercise features a simple geometric shape then this may be an adequate measure of the learner's ability to recreate the desired shape. However, when applied without modification to a carious tooth exercise, it introduces issues of construct-underrepresentation (Messick, 1995) because it does not account for the decision process or underlying rationale that led the learner to the produced cavity design so can result in them having a false impression as to their operative ability.

An often cited advantage of virtual reality simulation is that it reduces the need for direct tutor supervision (e.g. Luciano et al. (2009); Konukseven et al. (2010); Rhienmora et al. (2011)). However, without a tutor's presence, how are alternative approaches accommodated and what replaces the discussions that would normally occur between tutor and student in weighing up the conflicting factors? Without a tutor's guidance, will the student defer to the simulator's interpretation even when the approach they have chosen was perfectly valid? How might this impact upon the development of a reflective practitioner, able to reason about a clinical case and conclude the best approach when faced with a score telling them what percentage *wrong* they are? Does this suggest that an alternative approach is needed in place of absolute scores? Additionally, the simulator is only able to provide feedback on things it can actually measure in the virtual environment. This precludes correction on physical posture, handpiece grip, finger rest and many other material factors which are currently beyond the scope of the device's measurement, yet vitally important to early training.

Clearly, these devices have potential to make a positive contribution to the training of future dental practitioners, but these considerations mean that there is uncertainty as to their most appropriate usage. Their original intended position as a representational, digital equivalent, of the phantom head (San Diego et al., 2012) may have led to sub-optimal approaches being used which missed an opportunity to establish them with a distinct and complementary place in the dental curriculum.

1.1 Thesis structure

It was amongst this uncertainty that this project commenced with an initial broad research question of exploring "what is the optimal use of VR in dental education". In order to address this a review of contemporary uses of the technology in dental education and the identification of any gaps in the existing literature was undertaken using a Scoping Review methodology (**Chapter 2**). This review revealed questions relating to how learning can be measured, how motor skills are acquired and ultimately how learning in one context transfers to another; requiring a further focussed exploration of relevant literature (**Chapter 3**). This rigorous approach refined the understanding of the problem domain, leading to a refined statement of the research question (**Chapter 4**) to guide the remainder of the work.

The optimal usage of VR simulation in pre-clinical dental education as suggested by this earlier work is that it should focus on the acquisition of the cognitive knowledge that underpins operative skills and that this can be best achieved through the use of deconstructed part-task exercises. Therefore the remainder of work explored an approach to and the effectiveness of deconstructed learning in pre-clinical dental education using virtual reality simulation. This exploration began with the development of a novel approach to deconstructing operative tasks using task analysis in combination with VR and its application to an operative task (**Chapter 5**). The resultant Task Description is then explored through the lens of motor skills literature (**Chapter 6**) in order to guide the development of new VR exercises focussed on the underlying cognitive knowledge of the task (**Chapter 7**). The use of these exercises were compared to traditional teaching approaches with VR via a randomised control trial which explores quantitative measures of effectiveness and a thematic analysis of participant opinion of the deconstructed approach (**Chapter 8**).

As a result of the experimental work above, a recommended approach is proposed for future studies of a similar nature (**Chapter 9**) before discussing this work in light of developments in the use of VR that have taken place over this project's lifetime (**Chapter 10**). The work concludes with recommendations to simulator vendors in the development of future functionality and to dental educators faced with the challenge of optimising their usage of VR simulation for pre-clinical education (**Chapter 11**).

A more detailed overview of the organisation of this thesis is presented below:

- **Chapter 2: Literature Review** presents a review of the literature in the field of VR in Dental Education. It follows the Scoping Review approach (Arksey and O'Malley, 2005) and provides extended discussion of the wider context. Areas discussed include the simulation hardware, the realism of the simulation, feedback and assessment, and validity. The chapter concludes with a presentation of the gaps identified in the current literature.
- **Chapter 3: Measuring, Developing and Transferring Learning**, explores how motor skills learning is measured and how motor skills are acquired and demonstrated in other contexts. Insight from these wider fields of study are then applied to the use of VR simulation and how this informs VR's uses in the dental curriculum.

- **Chapter 4: The Research Question** presents the refined research question, aims and objectives for the remainder of this thesis as informed by the previous work.
- **Chapter 5: Task Analysis of a Dental Operative Task** presents a novel method for the task analysis of dental operative tasks, grounded in relevant literature, and then presents a study applying the approach to the analysis of a caries removal and cavity preparation task.
- **Chapter 6: Exploring the Task Description** explores the task description produced by the above analysis and relates it to the VR context.
- **Chapter 7: The Development of Deconstructed Part-Practice VR Simulation Exercises** describes the development of deconstructed and cognitive-focussed part-task exercises based on insight from the above Task Description.
- **Chapter 8: Comparing Deconstructed Learning with Traditional Teaching using VR Simulation** presents a randomised control trial where the effectiveness of the deconstructed task exercises are compared to a traditional teaching approach to teach the steps of establishing a caries free margin at the amelo-dentinal junction (ADJ) to cohorts of 1^{st} year dental students.
- **Chapter 9: An Approach for the Development of Deconstructed Exercises** describes a recommended approach based on the lessons learned from the above to guide future work of a similar nature.
- **Chapter 10: Discussion** revisits the hypotheses and reviews recent publications to discuss this work in the context of the latest literature published during the project's duration.
- **Chapter 11: Conclusion** summarises the conclusions from this work and makes recommendations for simulator vendors and dental educators for the future development of VR in Dental Education.

Chapter 2

Literature Review

2.1 Scoping Review

The use of Virtual Reality (VR) is a relatively new development in dental education, with the first commercial simulators appearing on the market around 2008 and experimental devices for research dating back as far as 1999 (Wang et al., 2016a). Research has been published on a number of areas of interest over this time so, in order to identify a suitable focus for this project a comprehensive review of the existing literature was undertaken. As a relatively young field, it was likely that there would be uneven coverage of the uses, effectiveness, advantages and disadvantages of these devices in the literature. Therefore, to assess which areas were most in need of further investigation a scoping review was conducted following the methodology described by Arksey and O'Malley (2005). This methodology provides a framework by which the extent, range and nature of research activity in a given field can be assessed and summarised so that any research gaps can be readily identified and disseminated in a comparatively short period of time compared with a full systematic review.

When introducing the scoping review methodology Arksey and O'Malley (2005) acknowledged that the approach does have its limitations, for example, whilst research gaps can be identified through omission, gaps through poor quality of research cannot be identified as the quality of the primary data is not assessed. Nor does it address how information gathered in this way is synthesised and relative weight is placed on different factors. However, the methodology allowed for a broad search of the literature that a more restrictive methodology might have excluded; permitting weighting to be placed on areas of interest in order to arrive at a suitable research question. The broad research question for this scoping review was:

What are the uses and applications of virtual reality in dental education?

A version of the following scoping review was published in the British Dental Journal (Towers et al., 2019).

2.2 Study Selection and Retrieval

Literature was sourced from both Web of Science and the Educational Resources Information Centre (ERIC) databases. Web of Science provides a single interface through which multiple literature databases can be searched simultaneously. ERIC is a smaller curated database focussed on educational literature sourced from journals, individual submissions and other sources of 'grey' literature.

To address the above research question, broad search criteria were selected to capture as much of the relevant literature as possible and searches were performed against the two databases. Results were then systematically filtered using defined inclusion and exclusion criteria. A summary of the process described below is shown in Figure 2.1.

Firstly, a title field search was selected to include all papers and proceedings etc. that mentioned terms relating to "virtual reality" as a concept. This included the terms "virtual reality", VR, "Virtual Environment" and "haptic*". 'Haptic' and 'haptics' are colloquial terms used in dentistry that have emerged as a shorthand term to refer to VR dental simulation so these were also included in the search. These terms were OR'd together and then AND'ed with the topic area of "Dent*". The wildcard being specified to capture variations from the stem including dental, dentistry, dentist etc.

More concisely, the search performed was:

> TITLE: (virtual reality OR Virtual Environment OR VR OR haptic*) AND
> TOPIC: (Dent*)

No other restrictions on study design, source or date of publication were applied, however, only papers written in the English language were included.

Searches using these terms were performed on both Web of Science and ERIC and results combined. This search could have been further refined to filter results by including terms such as "education" and "training", however, a manageable 128 results were returned so further filtering wasn't deemed to be necessary.

Next, a simple first pass of the titles was performed with the objective of identifying and discarding from further investigation any results that were clearly irrelevant. The criteria for exclusion of this stage were any results that were clearly not relevant and simply shared key words used in different contexts or found as a result of a greedy search

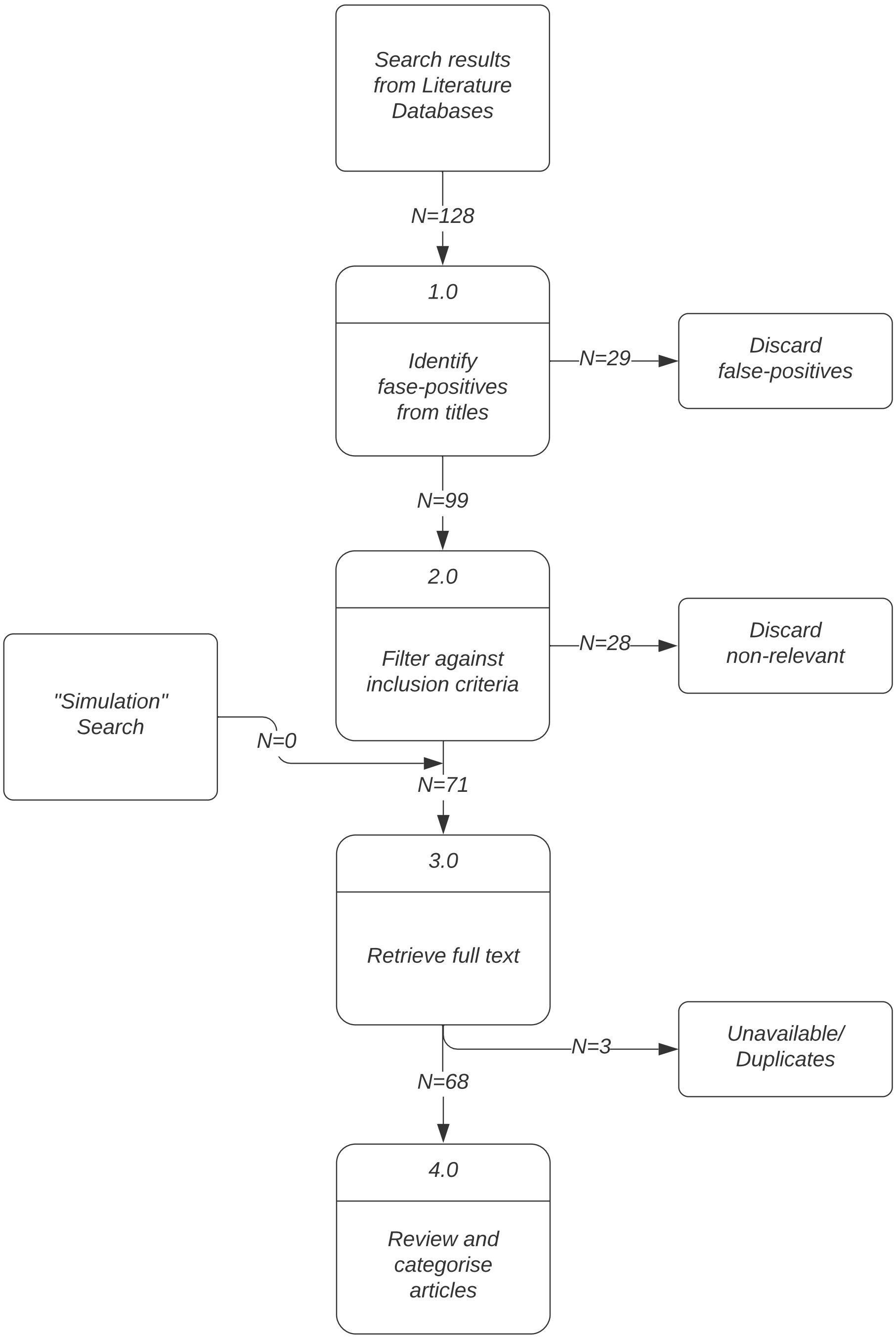


Figure 2.1: Literature search process

match. For example, the paper "Applying the Verona coding definitions of emotional sequences (VR-CoDES) in the dental context involving patients with complex communication needs: An exploratory study" was clearly not relevant but matched due to the presence of the terms VR and dental. During this stage, 29 search results were excluded.

For the remaining 99 results, abstracts were retrieved and reviewed in more depth. Inclusion criteria were defined and applied to the abstracts as follows:

- Papers describing the development of a VR dental simulator for use in education
- Papers evaluating the use of VR simulation in dental education and skills development
- Papers investigating the perception of VR simulation in dental education

Examples of the kinds of work that these criteria excluded include:

- Modelling dental occlusion via VR articulators
- The use of VR for dealing with dental anxiety
- The design of haptic algorithms/mathematical models to simulate cutting of tooth tissue in VR

The review of the abstracts against inclusion criteria eliminated a further 28 items from further study leaving the full text of 71 remaining papers to be retrieved.

Prior to retrieval of the full texts an omission from the search criteria above was addressed. The term "simulation" is a key term in this area, however, it presents difficulties in that it is used in a number of contexts (even when confined to dentistry); from mathematical models, to its various uses in role-play settings (e.g. simulated patient cases, simulated emergencies etc.), to its use for the simulated environments and interactions that take place in a petri-dish. So a separate search for this term was performed with results from the previous search excluded.

This new search produced 1,266 records, spread over 100 web of science categories. Clearly irrelevant categories were excluded, bringing the total down to 759. Then the same process as above was followed; first by using the titles to exclude obvious false positives (reducing the results to 69 candidate results) and then reviewing abstracts to include relevant results. After reviewing the abstracts, all relevant results had already been captured using the previous search approach, for example through existing terms such as referring to "virtual reality simulation" or "haptic simulation".

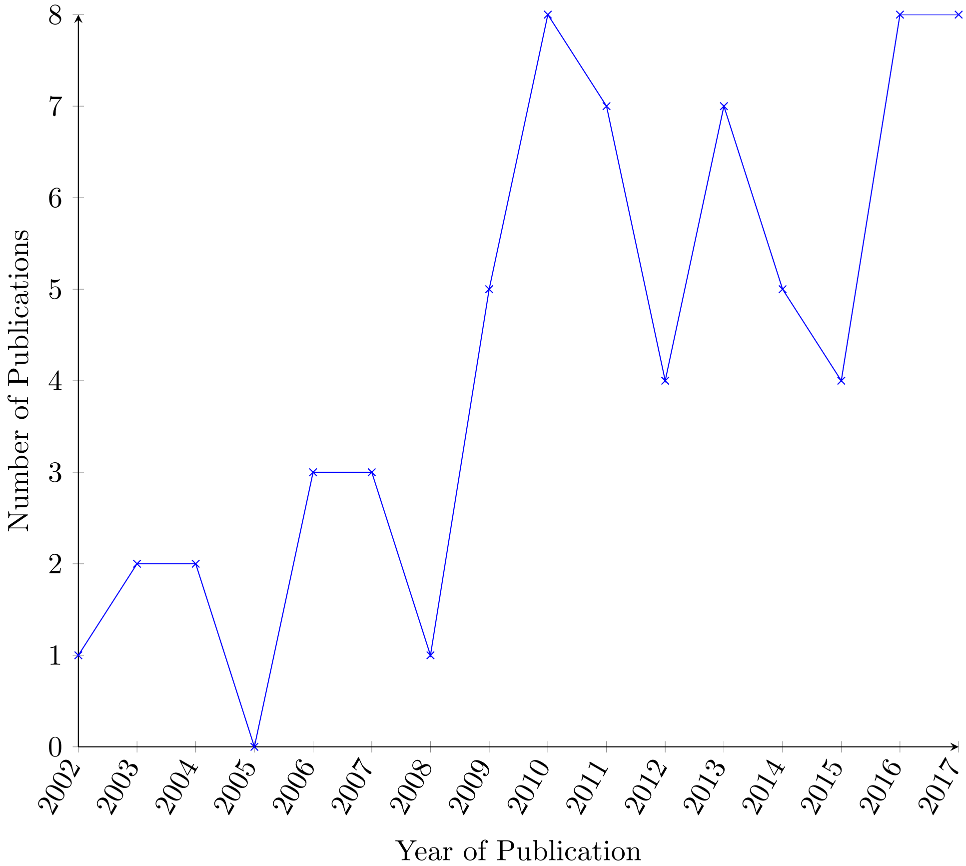
Finally, to ensure that any relevant papers published during the course of the review were captured, a weekly alert was set up using the same criteria to provide notification of any new publications. Results found this way were subjected to the same inclusion criteria as the original search results. A total of 8 papers were added to the original search in this way during the review.

The full text of the papers remaining at the end of the process were retrieved and read in full. Emergent themes were identified during reading and categorisation was performed and captured in a data extraction spreadsheet. Papers with shared themes were collated and discussed to present a narrative account of the existing literature and suggest areas in need of further investigation.

2.3 Results

Publication dates for the search results spanned from 2002 to 2017. A general upward trend in the number of publications can be seen over the time period as can be seen in Figure 2.2. with sharp increase in 2008. This date coincides with the first commercially available dental VR simulators so is likely to be as a result of more institutions acquiring the hardware and beginning to investigate their use in an educational setting.

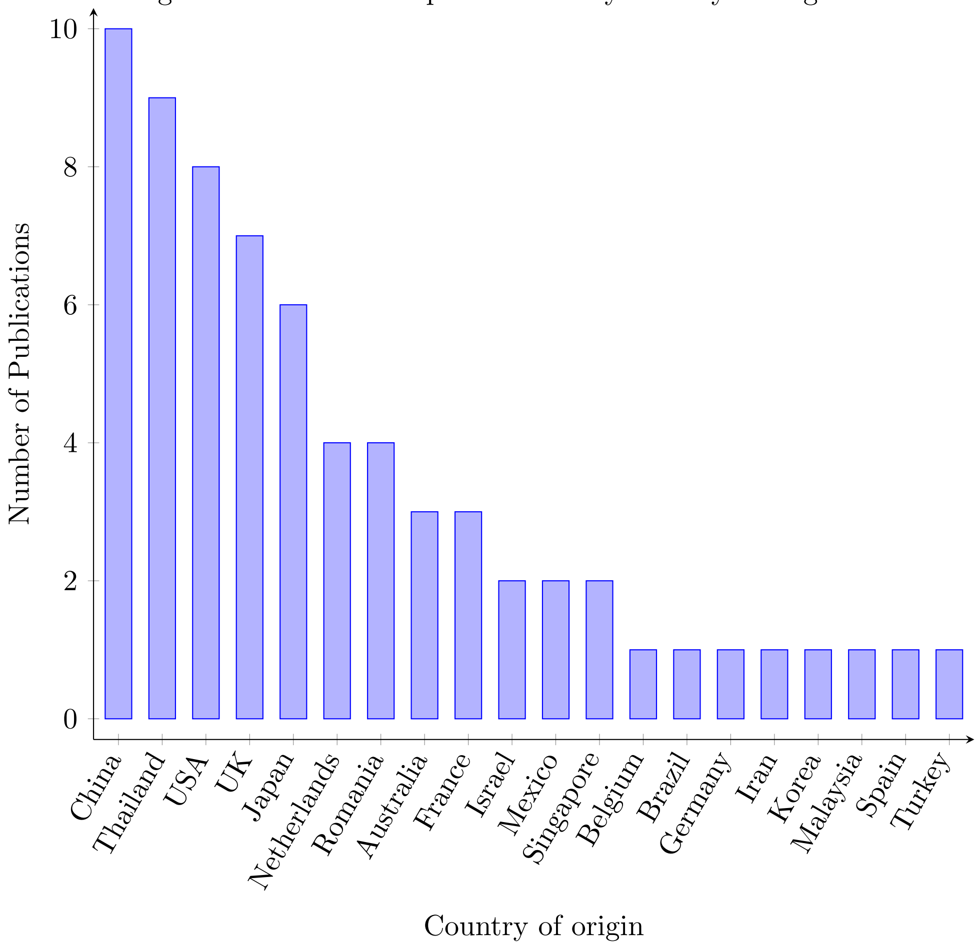
Figure 2.2: Publications per year



Comparing the origins of the identified works (based on the parent institution of the corresponding author) there is activity all over the world, but the two countries with the highest output in this field are China and Thailand. However, their contributions are dominated by a prolific research group in their respective country. The output from the next 3 highest countries show works from a wider range of authors.

The majority of the works surveyed used VR to simulate an operative or motor skill activity in an educational context. For most of these the task under investigation involved a cutting activity using a simulated dental bur. The acquisition of skills to precisely manipulate a dental handpiece to perform operative tasks is a core requirement for a dental student during their training so it is unsurprising that this is the focus of much of the work in this area. Table 2.1 shows a summary of the frequency of each task. It is interesting to note the subject specialism of journals in which research in this field has been published as over half of the works included in this review were published in computer science or technology enhanced learning publications. With the core function

Figure 2.3: Number of publications by country of origin



of these simulators being to improve on the acquisition of or to allow practice of dental clinical skills, the lack of a dental focus to these publications is surprising. Table 2.2 shows a summary of the broad categorisations of the publications in which these works were published. NB Both tables 2.1 and 2.2 were published in Towers et al. (2019).

Table 2.1: Operative tasks evaluated

Procedure or task	**Number of publications**	**References**
Cavity preparation	18	Mohamed and Luke (2002); Lian-Yi and Toshiyuki (2003); Wang et al. (2003); Kim and Park (2006); Wang et al. (2007); Wierinck et al. (2007); Wang et al. (2009); Rhienmora and Gajananan (2010); Wu et al. (2010); Konukseven et al. (2010); Gottlieb et al. (2011); Yoshida et al. (2011); Yamaguchi et al. (2013); Kozhevnikov et al. (2013); Eve et al. (2014); Wang et al. (2015b); Păvăloiu et al. (2016b); Plessas (2017)
Abstract shape preparation	12	Ben Gal et al. (2011); Urbankova and Engebretson (2011); Cormier et al. (2011); Urbankova et al. (2013); Kozhevnikov et al. (2013); Joseph et al. (2014); Ben-Gal et al. (2013); Eve et al. (2014); de Boer et al. (2016); Wang et al. (2016b); de Boer et al. (2017); Osnes and Keeling (2017)
Periodontal probing or scaling	6	Steinberg et al. (2007); Kolesnikov et al. (2009); Luciano et al. (2009); Wang et al. (2012); Deshpande et al. (2012); Yamaguchi et al. (2013)
Root canal treatment	5	Suebnukarn et al. (2010b,a, 2011, 2014); Toosi et al. (2014)
Implant placement	4	Kusumoto et al. (2006); Cormier et al. (2011); Syllebranque and Duriez (2010); Joseph et al. (2014)
Crown preparation	4	Suebnukarn et al. (2009); Rhienmora et al. (2009, 2010); Kikuchi et al. (2013)
Psychomotor tests	4	Kozhevnikov et al. (2013); Umemura (2014); Jamieson et al. (2015); Shahriari-Rad et al. (2017)
Caries diagnosis	2	Konukseven et al. (2010); Wang et al. (2016b)
Bone removal or surgical	2	Heiland et al. (2004); Ioannou et al. (2015)
Dental extraction	1	Wang et al. (2015a)
Cephalometry	1	Medellín-Castillo et al. (2016)
Endodontic measurements	1	Germans et al. (2008)
Local anaesthesia	1	Corrêa et al. (2017)

Table 2.2: Categorisation of publications

Journal Category	Number of publications	References
Computer science or technical publications (including eight from specialist simulation/ technology journals)	30	Mohamed and Luke (2002); Lian-Yi and Toshiyuki (2003); Wang et al. (2003); Luciano et al. (2004); Kim and Park (2006); Somrang and Chotikakamthorn (2006); Wang et al. (2007); Germans et al. (2008); Kolesnikov et al. (2009); Rhienmora et al. (2009); Luciano et al. (2009); Wang et al. (2009); Tse et al. (2010); Rhienmora and Gajananan (2010); Wu et al. (2010); Rhienmora et al. (2010); Syllebranque and Duriez (2010); Deshpande et al. (2012); Wang et al. (2012); Zheng et al. (2013); Jamieson et al. (2015); Ioannou et al. (2015); Wang et al. (2015a); Medellín-Castillo et al. (2016); Wang et al. (2016b); Escobar-Castillejos et al. (2016); Osnes and Keeling (2017); Plessas (2017); de Boer et al. (2017); Li et al. (2017)
Specialist dental education journals	15	Mohamed and Luke (2002); Steinberg et al. (2007); Wierinck et al. (2007); Konukseven et al. (2010); Suebnukarn et al. (2010b); Gottlieb et al. (2011); Ben Gal et al. (2011); Urbankova and Engebretson (2011); Boer et al. (2013); Ben-Gal et al. (2013); Kikuchi et al. (2013); Urbankova et al. (2013); Eve et al. (2014); Suebnukarn et al. (2014); de Boer et al. (2016); Wang et al. (2016a)
General dental or medical journals unrelated to simulation or learning	8	Heiland et al. (2004); Kusumoto et al. (2006); Suebnukarn et al. (2009, 2010a, 2011); Joseph et al. (2014); Roy et al. (2017); Corrêa et al. (2017)
Technology enhanced learning journals	8	Bogdan et al. (2011); Cormier et al. (2011); San Diego et al. (2012); Păvăloiu et al. (2016b); Fernández et al. (2016); Păvăloiu et al. (2016a); Sabalic and Schoener (2017); Shahriari-Rad et al. (2017)
Materials journals	5	Steinberg et al. (2007); Luciano et al. (2009); Sofronia et al. (2012); Toosi et al. (2014); Wang et al. (2015b)
Psychology journals	1	Yamaguchi et al. (2013)
Open journals without specialisation	1	Umemura (2014)

From the tagging and categorisation process, a number of themes emerged from the literature:

- The simulation hardware
- The realism of the simulation
- Automated feedback and scoring
- Validation of the exercises and the role of the tutor

The remainder of this review is structured around these thematic areas and will address each of them in turn.

2.4 The simulation hardware

The hardware components of the simulator define the environment that the user will interact with. Decisions made when choosing hardware components, their physical layout and whether or not any aspect is included as a feature of the device can impact upon the interaction affordances and will clearly have consequences for the range of uses the simulator is appropriate for.

2.4.1 Form factor

A number of different physical simulator configurations which claim to provide a 'Virtual Reality' experience are reported in the literature. There are no formalised requirements for the features and specifications a VR dental simulator should conform to, so a significant variance can be seen between the simulators used across the literature. However, they can be grouped into four broad classifications with an increasing likeness to the dental operative environment.

Desktop PC

The hardware classification most removed from the ergonomics of the real operative environment, as used by one paper in this review, was the desktop PC. Here, a conventional keyboard and mouse was used to control an on-screen handpiece to prepare a cavity in a simulated tooth (Mohamed and Luke, 2002). The system reported did not make any concessions to the kinds of physical interactions that take place when carrying out this procedure in reality.

Figure 2.4: 3D Systems Touch X. Reproduced with permission from 3D Systems, USA

Haptic Desktop

A conventional computer mouse is only able to represent x,y movements made in 2D space. This is insufficient to replicate the 3D interactions that occur in a real-world dental operative environment. To overcome this, a number of groups replaced the mouse as the primary interaction device with one or more haptic arms which allow the user to move a pointer in 3 dimensions (Figure 2.4). Many haptic arms are also able to provide a positional reading for the attached stylus' rotational orientation giving a full 6 directions of freedom (6-DoF) making them suitable for modelling the location and orientation of a dental tool in 3D space. Additionally, a haptic arm is also able to provide tactile, or "haptic", feedback in some of these directions of movement such that a user is able to feel like they are touching an object when they come into contact with it in the simulation. These sensations are created by electric motors at the articulation points of the haptic arm which resist a users movements and by applying varied levels of torque, the variable densities of different materials can also be simulated.

These 'haptic desktops' (Figure 2.5) improve upon the Desktop PC simulators by allowing 3D movement of the dental tools and tactile feedback, however they do not attempt to recreate the ergonomics of the dental operative environment. Haptic Desktops were the most common simulator configuration found in this study as reported by 28 of

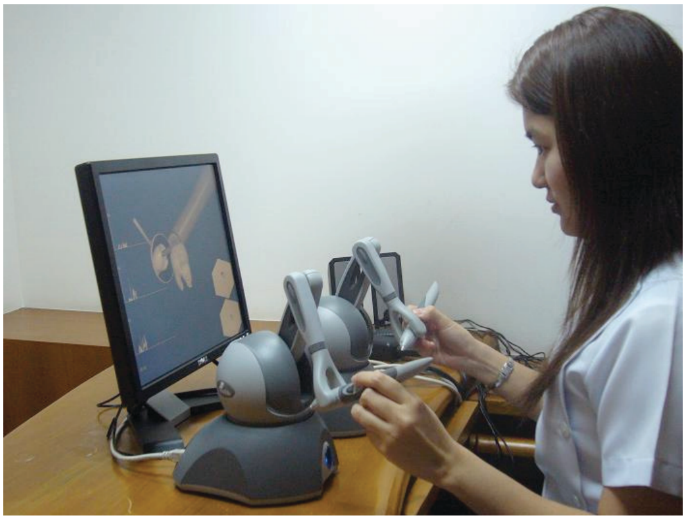


Figure 2.5: A student using a haptic desktop. Reproduced with permission from Siriwan Suebnukarn

the 69 works in this review.

Dental Skills Trainer

Dental skills trainers attempt to address some of the ergonomic issues of haptic desktops. These systems (e.g. Figures 2.7 and 2.6) add a bespoke enclosure to the computer and build in dental-specific features such as height adjustment to achieve an ergonomic working posture, a finger rest stage, foot pedals to operate the handpiece and replace the default haptic arm stylus with a simulated handpiece. Additionally, this configuration positions the haptic arm in a location so that the tool is closer to where it would be in the operative environment. Dental skills trainers were reported as the hardware used by 17 of the studies in this review.

Figure 2.6: The Simodont dental skills trainer. Reproduced with permission from Moog, USA

Figure 2.7: The Virteasy dental skills trainer. Reproduced with permission from HRV Simulation, France

Digital Phantom Heads

Finally, four publications reported the use of a digitally-enhanced phantom head. Here a student uses a traditional phantom head configuration and cuts plastic typodont teeth using a real handpiece. The handpiece's state and movements are captured by 3D cameras and, because the dimensions and position of the typodont tooth are known, it allows a digital recreation of the material removed to be represented in software and be subjected to analysis to provide automated performance feedback. It could be argued that these systems are not 'Virtual Reality' in a traditional sense as the interactions are taking place in the real world and only retrospectively digitally recreated to enable computer assisted assessment and performance capture.

Simulator form factor

These four broad classifications show significant differences in the designs of the simulator hardware on which the literature of this field is based. There are no standards for evaluating or comparing dental simulators (Wang et al., 2016a) so it is unclear what the implications of this variance will be in terms of both evaluating research contributions or the impact of the use of VR simulators in a dental education context.

Even the dental skills trainers with their adaptations to recreate some of the physical environment of the real operating environment are still very far removed from the ergonomics present in the clinical environment. Differences between the virtual procedures and the real one can create difficulties and change habits (Escobar-Castillejos et al., 2016) and as can be seen with the haptic desktop configuration, the difference to the real operative environment can be significant. This poses the question to what extent can the physical simulation environment deviate from its real world equivalent before the experience is devalued or even becomes negative, requiring skills to be un-learned?

These variances are compounded by omissions in the simulator design. Tse et al. (2010) observed that their simulator lacked cheeks and gums, and that this allowed a user to position the handpiece at angles that would not be possible in the oral cavity. Their group planned to add rubber cheeks to constrict the range of available movement so that 'bad habits' were not developed by becoming accustomed to techniques that don't work in practice. However, this desire to add additional realism to the physical environment must be weighed against the physical considerations of the haptic arm that provides the force feedback. This *arm* structure is not present in a real operative environment, so collisions could occur in the simulator between the arm and the rubber soft tissue which would not happen in the real situation. Additionally, whilst the rubber cheeks would physically constrict the range of motion, the position of these cheeks might not be represented in the system's graphic representation, creating a disagreement between the tactile and visual aspects of the simulation. These factors are illustrative of some of the trade offs and compromises that are necessary in the design of VR simulators, so it is important that these decisions are informed by careful investigation to avoid unexpected issues.

The degree to which visual and haptic sensory inputs spatially agree is known as their "co-location" and it is important for increasing the user's sense of presence in a virtual environment (Dinh et al., 1999). In many of the haptic desktop simulators there is a poor co-location between the position of the haptic arms and the location of the visual display of the virtual world, and whilst this does not appear to have impacted upon the perceptions of the usefulness of the simulators (Suebnukarn et al., 2009), poor co-location has been shown to be an important factor and greatly impact on user performance (Swapp et al., 2006). An innovative step to improve co-location in virtual worlds was taken by Rhienmora and Gajananan (2010) where their earlier work was adapted to use a augmented reality (AR) system. This projected a tooth on to an AR marker making it appear as if it was sat on the user's desk and could be interacted with via a table-top haptic arm. The AR view also projected a handpiece over the haptic arm's default stylus, giving the visual appearance of holding a real handpiece. Importantly, the close co-location and the AR view provided the additional advantage of the user being able to see their own hand and ensure they were holding the handpiece correctly.

Other than the AR representation in Rhienmora and Gajananan (2010), the remaining simulators in this review prevented the user from seeing their own hand; either by being obscured by the computer display equipment or as a result of the user's gaze being directed away from the handpiece to view their performance on a computer monitor. As observed by Xia et al. (2013), this may impact on eye-hand co-ordination and lead to difficulties in translating the skills learned in the VR environment to a real life context. Research has shown that the human brain continually uses visual information from the hand when making movements (Saunders and Knill, 2003) and that different areas of the brain are triggered when observing a real hand and a virtual reality hand (Perani et al., 2001). That Perani et al. (2001) showed the virtual reality representation of a hand did not access the full motor knowledge available to the central nervous system suggests that the impact of not being able to perceive an operator's real hand should be further investigated.

The importance of the physical simulator design and the way in which the user interacts with it has been recognised in the aviation simulation field and codified into certification specifications. The European Aviation Safety Agency provides categorisations for different kinds of simulator (EASA, 2012). Flight simulation training devices are divided into categories from a Full Flight Simulator (FFS) where a full size replica of a specific aircraft model is created with views out of the cockpit and full simulation of the movements and forces exerted by the aircraft - to a Basic Instrument Training Device (BITD), which represents a basic student pilot station for a class of aircraft. "Procedural" aspects of aviation can be adequately trained and assessed on BITDs but are recognised as insufficient for other aspects. If training is undertaken on an appropriately advanced category of simulator, this training time can be counted towards minimum flying time requirements and shows the significance that the aviation industry has placed on the accuracy and fidelity of the simulation environment (EASA, 2011).

2.4.2 Simulated tool

The device attached to the haptic arm and held by the user can be either be a generic stylus, or can be replaced with a facsimile of the real dental instrument. Most works in this survey (40 of the 65 who reported using a haptic arm) opted to use the default stylus as provided by the haptic arm manufacturer. As can be seen in Figure 2.4, this device resembles a marker pen, has a rubberised grip and includes a button for affecting actions within the software. Only 11 of the works surveyed reported replacing the stylus with a realistic facsimile of the relevant dental instrument.

There is very little discussion within the literature which considers the impact of interacting with the simulation using a physical device that does not resemble the real instrument. In the development of a periodontal simulator, Luciano et al. (2009) were driven by a desire to create a realistic simulation. Whilst going to some length to stress the importance of the haptic feedback, even going so far as to call it "unquestionably critical", their test apparatus used the stylus from the haptic arm rather than the (much smaller) periodontal probe. Similarly, Toosi et al. (2014) presents a simulator that has the ability to simulate the filing of the canals using virtual K-Files. Again, the on-screen files are controlled by the haptic arm's stylus which is a very different shape and requires a completely different grip to what would be required to use a real K-File. Only Corrêa et al. (2017) questioned users on their preference for the shape of the interaction device (a simulated syringe or the default haptic stylus) and found results differed based on the participant's level of experience. Some beginners expressed a preference for the haptic stylus, but more experienced practitioners strongly preferred the syringe. However, whilst ascertaining a preference is commendable, user preference is of secondary concern to the reliability of skills developed when using the simulator as an educational tool.

The behaviour of the device when in operation is also of interest. Tse et al. (2010) noted that the inertia and weight of the haptic device were too high and unlike Konukseven et al. (2010) and Somrang and Chotikakamthorn (2006) was unable to recreate the handpiece's vibration when active. Konukseven et al. (2010) reported that the vibration should be at 500hz with an amplitude of 0.3N and observed that it has a considerable effect on the perceived realism of the haptic sensations. As Tse et al. (2010) acknowledged, the impact of not simulating the vibration of the handpiece is unknown, however, given its importance for perceived realism, the impact its presence or absence has on the acquisition and transferability of skills should be investigated.

Five studies reported abstract modes of interaction within the simulation environment that differed notably from the real world equivalent. Most took advantage of the button on the haptic arm's stylus to add behaviour, for example, to:

- control the handpiece motor for a cavity preparation exercise (Suebnukarn et al., 2009),
- trigger the 'grip' of virtual forceps in an exodontia exercise - rather than making a gripping motion by opening and closing physically simulated forceps (Wang et al., 2015a),
- place the handpiece in 'restoration mode' where the restorative material *grows* from the tip of the handpiece to fill a prepared cavity instead of using an amalgam carrier (Kim and Park, 2006).

It is unexplored what the impact of an inaccurately modelled interaction device has on the transferability of the skills, the performance of the operator, or the face validity of the simulation. Only Corrêa et al. (2017) recognised that the stylus has the affordance of a pen-like grip and that this might be detrimental to learning correct technique. It has been shown that holding a dental handpiece incorrectly has a significant impact on performance (Makinson and Hume, 1982), so an inaccurately modelled device is likely to impair the ability to achieve a correct grip and negatively impact performance.

However, whilst it is easy to dismiss these non-standard interaction modalities due to their lack of similarity to the real situation; this should not prevent deliberate modification or innovations within the environment. The underlying aim of these simulators is to provide an educational benefit and a computer simulation doesn't necessarily have to precisely model the laws and norms of the real world if diverging from them leads to interaction modes which aid understanding or accelerate learning.

2.4.3 Force reproducibility

As the user's primary point of interacting with the simulation environment, the haptic arm's capabilities are an important area to be considered. The following subsections discuss factors relating to the ability of the haptic arm to provide appropriate force feedback.

Degrees of freedom of force feedback

As discussed above, haptic arms reproduce the sensations of touching a virtual object through the force exerted by electric motors at the articulation points. To fully represent the position of a dental tool, a simulator's haptic arm must be able to track all of the possible movements, or degrees of freedom. These include the translational: x, y and z

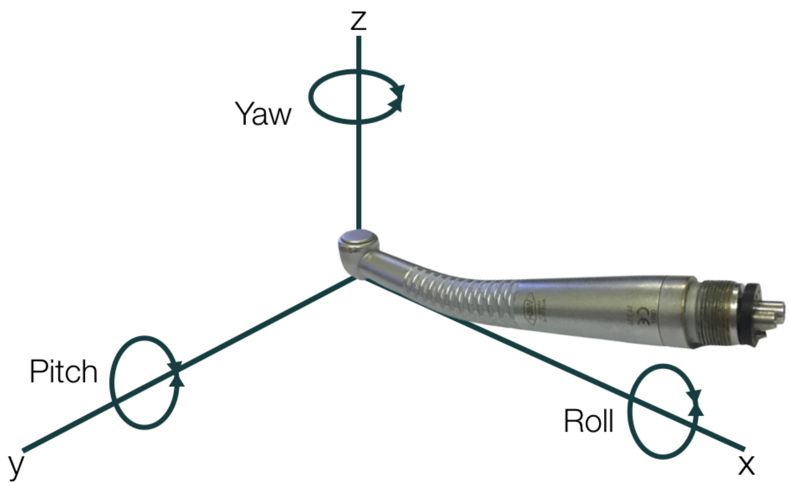


Figure 2.8: Axes of movement

planes; and the rotational: pitch (vertical rotation), yaw (horizontal turning) and roll (tilting rotation) as shown in Figure 2.8. As recognised by both Wang et al. (2016a) and Escobar-Castillejos et al. (2016) there are limitations in the existing hardware for the provision of haptic feedback. Most simulators in this review were equipped with haptic arms that can only provide haptic feedback and resistance against translational movements (a 3 Degrees of Freedom of Force Feedback (3-DoFFF) haptic arm). To provide resistance against rotational movements a 6-DoFFF haptic arm is required. Unfortunately, at present these are cost-prohibitive and consequently 46 of the studies in this review reported using 3-DoFFF haptic arms.

A consequence of using a 3-DoFFF device when simulating a dental procedure is that the appropriate haptic feedback can not be provided when the user makes a rotational movement. Wang et al. (2012) recognised that this permits the tool to rotate in ways that are not possible in reality, resulting users being able to cheat when measuring periodontal pocket depths by rotating the probe *through* the gingiva and viewing the markings on the tool more easily. As a result they recommended preventing this by moving to a 6-DoFFF device so that multi-point contacts could be modelled.

The lack of multi-point modelling also introduces difficulties when attempting to successfully carry out some simulated procedures. Multi-point contacts allow the user to judge the flatness of a surface by eliciting a rocking movement or guide the tool preventing erroneous rotation. In the absence of a 6-DoFFF device, the simulator will either have to produce an un-natural translational corrective response or, as Wang et al. (2012) found, allow the angle of the haptic arm to correspond with a position that would be impossible in the real world. Only 4 studies in this review reported the use of a 6-DoFFF device.

Strength of force feedback

The 3D Systems Touch and Touch X haptic arms, as used by many of the studies in this review, are capable of outputting a maximum of 2.35 N/mm of force, which is reported as being lower than the contact stiffness of enamel (Wang et al., 2016a). Xia et al. (2013) cautions against the use of unrealistic haptic feedback during training as it can mislead operators when carrying out real surgery. However, whilst this maximum of 2.35N/mm is lower than the contact stiffness of enamel, this is an absolute value, and may not represent the contact stiffness of enamel when that contact is with a spinning bur in a high speed handpiece. So, calibration of the force feedback should be undertaken so that the forces can be appropriately modelled (Escobar-Castillejos et al., 2016). In some cases, the discrepancy between the haptic arm's maximum output and the force required in the real procedure can be by several orders of magnitude. For example, in Wang et al. (2015a) a simulation of the tooth extraction process was presented, however, the haptic arm used was only capable of a maximum output of 3.3N/mm. Whilst the system is described as providing a heuristic model of the process rather than a physical simulation, it is clearly significantly lower than the measured forces involved in the tooth extraction process which can peak at 199N/mm (Cicciù et al., 2013). Just 5 of the works in this review directly mentioned that there may be issues with the maximum output force with a further 8 inferring an awareness of the issue.

To date, only de Boer et al. (2017) has investigated the use of haptic feedback on the performance and satisfaction of dental students. However, this study only dealt in the absolutes of the provision of force feedback or its absence. The impact of unrealistic degrees of force feedback between these extremes was not considered. It is worth noting that plastic typodont produce a different haptic sensation to a real tooth when cut with a handpiece, but these differences do not preclude them being a useful substrate on which dental skills can be learned. No research that investigated the implications of training with an unrealistic level of force or only a subset of the forces involved in the real procedure could be found.

2.4.4 Finger rests

A finger rest is a critical element for the safe operation of a dental handpiece. It provides stability, a fulcrum for the actions of tools and maintains contact with the patient so that unexpected movements can be rapidly detected to avoid inflicting injuries. Its value is stressed to students in recommended curricula (Field et al., 2017, 2018) and recommended operative text books (e.g. Banerjee and Watson (2015); Field (2016)).

Despite this, only 14 of the works surveyed reported or illustrated a platform where a safe operating finger rest could be achieved, and only 6 explicitly recognised that its provision was a hardware design consideration. An additional 5 works reported that the wrist rest from a mouse-pad was available. However, 22 papers reported hardware that lacked any finger rest at all, resulting in the operator's hands being completely unsupported. Even those simulators that do provide a structure on which to achieve a finger rest offer structures that are very different from those encountered when working with real patients (Wang et al., 2016a).

The variable importance placed on the finger rests provided by haptic systems is illustrated by the existence of a 'hand stability system' (Deshpande et al., 2012). Here, a misconception existed, that poor handpiece control was a result of the operator's nervousness. The paper reports development of a device that would, with Velcro straps and a hand grip, minimise a user's shaking. Whilst this device did result in an improvement in performance, it prevented the user from correctly holding the handpiece and achieved stability by simply constraining their movements. It may be the case that a correct handpiece grip, and the presence of a position to achieve a finger rest, would have avoided the perceived problem altogether.

Wang et al. (2015a) argues the importance of "what you see is what you feel" for the immersion and interaction in VR. However, throughout the literature the importance of this similarity is reserved for the feel of the bur as it cuts through simulated material and ignores the wider physical simulation environment. The inability to physically feel some of the items visible in the virtual world, and the lack of co-location between the user and the supporting structures where, in the real environment, they would achieve a safe finger rest is unknown. This lack of appreciation of the importance of the finger rest and its consequent absence may have impacted on the results of many of these studies, so their reliability as a basis for the generalisation of wider operative skills acquisition may be questioned.

2.4.5 Simulated 3D depth

Dentistry is a spatially complex task occurring within the confined area of the mouth. A dental student must learn to operate dental tools and make precise movements within the oral cavity to provide operative treatments; sometimes needing to observe their motions in reverse through a mirror. The ability to perceive and reason about 3D space comes from our ability to see in stereopsis. This arrises from a number of factors: Binocular disparity hypothesis - that stereopsis is as a result of binocular vision from having two eyes; Visual parallax hypothesis - that stereopsis results from the perception of different views of the same scene over time or through movement; Cue coherence/depth magnitude hypothesis - that the perception of depth is related to perception/interpretation of cues such as relative size, occlusion, shadows etc. (Vishwanath, 2014).

Computer monitors are conventionally designed to display images in 2D. However, some models are capable of simulating stereoscopic 3D by presenting a different image to each eye in conjunction with an appropriate pair of 3D glasses. This effect is achieved in one of two ways, by either: quickly alternating between an image for the left eye and right eye, with a pair of synchronised active 3D glasses blanking out the view for the eye the image is not intended for (Time-Multiplexed Stereoscopic Display) or, by displaying both images simultaneously using alternate rows of the display for the left and right eye images passed through rows of alternately orientated polarising filters (Polarization-Interlaced Stereoscopic Display) with a pair of polarised glasses which only allow the light for the correct eye to pass through (Geng, 2013). 3D displays of these kinds were the most common found in this review and were reported by 22 of the papers.

Very little research has investigated the importance of 3D in VR for dental education. de Boer et al. (2016) claimed that adding stereoscopic 3D to their simulation resulted in significantly better exercise performance, and when questioned, was preferred by 90% of participants to the 2D equivalent. However, this study failed to account for a confounding factor whereby users of the 2D display reported discomfort in the form of eye strain and headaches caused by still being required to wear the 3D glasses. The authors correctly observe that both sets of participants should wear glasses to isolate any effect that this may have, but the 2D users could have been provided with non-polarising "dummy" glasses to avoid the discomfort. The presence of visual discomfort in one group of participants could mean that the preference towards the 3D display could be considered unreliable.

To further investigate the importance of stereopsis in a dental context, Al-Saud et al. (2017) compared performance in 2D and 3D for the occurrence of depth related errors in the execution of a cutting task. This work showed a significant difference in performance, and that without the 3D display students were more likely to make errors that are associated with issues related to depth perception. However, as conceded by the authors, this and the de Boer et al. (2016) study above only consider the impact of stereoscopic 3D in isolation. Providing a 3D effect that only caters towards binocular disparity is likely to be insufficient for such a spatially complex task as operative dentistry. It has been shown that depth judgement is improved when a stereoscopic 3D is combined with parallax head movements (Dees, 1966) and the movement of the head to check for correct angulation of a tool relative to the tooth surface is a common motion and part of the technique taught in pre-clinical skills courses.

Only seven papers reported simulators that were able to track the user's head movement to allow for visual parallax depth to be perceived. Of these, 6 reported the use of fully immersive VR and AR headsets which provide an independent display for each eye facilitating binocular vision and contain gyroscopes to track the movement of the head which is reflected with a corresponding change to the view seen by the user. The remaining paper used a stereoscopic monitor and tracked the user's head position with a camera (Tse et al., 2010). The authors reported that the lighting conditions resulted in the users moving their heads slower than desired, but interestingly, used this time to plan their next task. It is important to establish if head tracking is a requirement or if proxies such as the provision of additional on-screen viewpoints would be adequate (or even preferable) during training.

The complete absence of 3D has been shown to have a negative performance impact so this may question the reliability of the findings of 17 papers in this review that did not provide any 3D view. However, it is unknown what the impact of not being able perceive different forms of 3D is or how this affects transfer to the real context where students will not only be able to perceive these additional depth cues but will also be encouraged to use them to improve their performance.

2.4.6 Overview

It is clear that there are many differences between the hardware of the systems used to document the use of VR as a tool for training dental operative skills. These differences mean that it is unclear if an educational task validated on one device is translatable to a different system. There is also a highly visible difference between the simulators and the ergonomics of the real operative situation, even down to the very way in which it is perceived. Therefore, additional work is required to establish the impact of these differences to the transferability of the skills developed in a VR simulated context.

When compared with the importance that aviation places on the simulation environment, some dental simulators only present a superficial likeness to the real world environment, so it could be questioned if dentistry is being 'simulated' (in the strictest sense) at all? If dentistry is not being *simulated*, then how can these devices, with all their potential, be put the best use to serve dental students and institutions?

2.5 Realism

The realism of the simulation and the need for the simulation to be as realistic as possible is a recurring theme throughout almost half of the studies in this review. 20 papers directly claimed it is essential and a further 14 clearly agree with this position. Phrases such as "resemble reality as closely as possible'" (de Boer et al., 2017) and "the system should simulate as closely as possible the real clinical activity with patients" (San Diego et al., 2012) were common. This theme can also be seen in the wider dental education literature, for example, where tutors often attend to the differences between the phantom head and a real patient during teaching sessions (Hindmarsh et al., 2014). Such is the strength of this desire for realism that it was claimed that the use of extracted teeth is the optimal substrate for training on (Boer et al., 2013) but practicing on real patients is a much more effective way to train than using traditional methods (Wang et al., 2016a) (although they did concede the undesirable risks this would pose to the patients). However, in no place where these claims of importance are made was there reference to any relevant supporting literature.

The underlying assumption is that simulation needs to be as realistic as possible and a lack of realism limits its value (Ruthenbeck and Reynolds, 2015), so by extension, the more realistic the simulation is, the more effective it is. In establishing realism, the extent of *what* is simulated should be considered. It is claimed that simulation of the entire oral cavity including all teeth and soft tissues is an “indispensable” requirement to train fine motor skills in the oral cavity (Wang et al., 2016a) or that even the wider body should be simulated to allow freedom to practice any procedure that is wished (Ruthenbeck and Reynolds, 2015). However, the costs and computational overhead of simulating to this extent are likely to be prohibitive. There is also an internal disagreement, that on the one hand, simulation has the advantage to focus on individual aspects of treatment (Ruthenbeck and Reynolds, 2015), but on the other hand, the wider body (perhaps beyond the scope of the treatment being attempted) should be included. Including more factors in the simulation requires these to be accommodated by the learner, potentially reducing their ability to focus on the improvement of the task in hand.

2.5.1 Developing a simulated tooth

Roughly $\frac{2}{3}$ of the papers in this review (46 of 68) reported the use of realistic teeth or oral structures in their simulation. The teeth for use in VR simulations have been created using: generic anatomically accurate 3D models (e.g Sofronia et al. (2012); Kolesnikov et al. (2009)) or by using cone beam computer tomography (CBCT) (e.g. Boer et al. (2013); Syllebranque and Duriez (2010)). In the case of the generic models, these can be created using 3D editor software or purchased from online asset stores (Păvăloiu et al., 2016a). For CBCT sourced models, using the process described by Boer et al. (2013): a CBCT scan of a real patient is taken, and the ‘Digital Imaging and Communications in Medicine’ (DICOM) data is processed through software to extract a 3D model. This model is then split based on the radio-translucency of each structure so that the crown, dentine, pulp and any carious lesion can be separately modelled and assigned different colour and density values.

Once the physical shape of the tooth has been established and it has been segmented into the different types of tissue, the hardness of these structures must be modelled by the simulator software so that these differing densities can be physically recreated by the haptic arm. The acquisition of appropriate hardness values has been achieved by both empirical definition (Yamaguchi et al., 2013) and experimental (Yoshida et al., 2011). However, in both of these studies a simplification to the resultant tooth models was made so it only represented ‘hard tissue’ and ‘pulp’, i.e. the dentine was not distinguished and was modelled as a continuation of the enamel. In Boer et al. (2013) all materials in the tooth models are modelled fully, however it does not state how the density values for their simulator were arrived at (simply that they are aligned to a greyscale value from the CBCT scan). Boer et al. (2013) asks a pertinent question of if the teeth are sufficiently realistic for use in dental education, and Wang et al. (2012) goes further by questioning how realistic surgical simulation needs to be and what the required level of fidelity is?

Finally, to simulate caries within the tooth, the tactile feel of caries must also be modelled by the simulator. Wang et al. (2016a) attempted to calibrate the simulated stiffness of mild and severe caries in a simulated tooth by asking 5 dentists to probe regions containing various levels of demineralisation. However, they found substantial variance between participants in the preferred levels of stiffness delivered by the haptic arm for each predefined level of demineralisation. Similarly, Konukseven et al. (2010) used density values derived from micro-mechanical measurement, but permitted participants to adjust the values and 'fine tune' the density settings to receive what they considered a realistic feel. Both of these studies featured multiple qualified dental practitioners, yet differing views were given when asked what the appropriate hardness values for different dental materials should be. This suggests that there may be an element of subjectivity in the perception of the forces involved. When operating on real tissue, the density and feel of each material is unquestionable. For example, if the operator is cutting through enamel, it is evidently enamel - despite any naturally occurring variations in the density found in enamel (Weidmann et al., 1967) it is, unquestionably, enamel. However, when operating on a simulated tooth, if the density being reproduced by the simulator does not match the pre-existing mental model held by the operator, then the simulator is likely to be considered *wrong*. This judgement would be regardless of if the represented density falls within the envelope of real-world values because the simulator must always correspond to the subjective representation of the simulated world in the user's head (Wages et al., 2004).

These differences in the perception of the simulated tissue that have been uncovered whilst attempting to develop teeth for use in simulation point to a deeper relationship between the simulated representation and how it is perceived which may have consequences for the use of VR in dental education.

2.5.2 The Treachery of Simulations

Whilst a thorough examination of the philosophy of art is beyond the scope of this thesis, the implications for realistic representations have been discussed by artists, illustrators and more recently by roboticists. Through these lenses, the realistic representation of a tooth and the pursuit of ever more realistic simulations may lead to undesirable consequences.

Figure 2.9: "The Treachery of Images" by René Magritte, 1929, Oil on Canvas

In his famous painting "The Treachery of Images" (Figure 2.9), René Magritte shows a pipe in profile with the caption "ceci n'est pas une pipe", which translates from the French as "this is not a pipe". This challenges the viewer with a contradiction; there is a realistic pipe, but the caption reminds the viewer that it is not in fact a pipe but merely a representation. As Magritte himself remarked, "The famous pipe - how people reproached me for it? And yet, could you stuff my pipe? No, it's just a representation, is it not? So, if I had written on my picture 'This is a pipe', I'd have been lying!" (Torczyner, 1977). Equally, a tooth appearing in a simulated environment is not a tooth. As discussed above, it is an interactive representation of a 3D model, created from inferred values of the radio-translucence of a scan of a tooth, but it is not a tooth. However, unlike Magritte's painting, it is possible to 'stuff the pipe' because it is represented in an interactive computer simulation with simulated dental tools. So, visually it resembles a tooth, it can be interacted with as if it were a tooth yet, there remains the contradiction that "ceci n'est pas une dent".

Writers in illustration have discussed how an image representation is merely an 'icon' or a placeholder to represent the idea of an object. Just as a simulated tooth is not a tooth, the word 'tooth' written here using the letters of the ISO Standard Latin Alphabet is not a tooth either. Both conjure the idea of a tooth, which can vary from person to person, but in themselves, neither icon is *actually* a tooth. But, as abstract iconographic representations begin to resemble a lifelike visualisation of the object concerned, the idea *becomes* the object and the individual flexibility of what that icon represents is compromised (McCloud, 1994). If this same principle can be found with dental simulation, then as the visual realism of the tooth in the simulator increases, the simulated tooth will cease to be an abstract substrate on which a skill can be learned and practiced and shift to being a 'specific' tooth and be expected to share all of the the properties, feel and anatomical cues of the real structure.

This idea can be taken further. In robotics, the concept of an "Uncanny Valley" (Mori, 1970) has been proposed. This theorises that as the human-likeness of a robot increases there is a "dip" in our acceptance of how comfortable we are with their appearance. For example, a metallic, sharp angled, mechanical device would clearly be robot and would be accepted as such with no issues. However, as the appearance transitions to an increasingly more human-like aesthetic, a point is encountered *just* before it is indistinguishable from a real person when the acceptance sharply reduces. In this Uncanny Valley, attention flips from the (perhaps hundreds of) factors that make the robot seem realistic to focus on the "un-human" aspects. The icon becomes an individual, but one who lacks a certain spark, or the other subtle factors of "familiarity" found in a healthy living being. Perhaps, similarly, by striving to create more and more realistic representations of a tooth the slight imperfections and differences from the real thing become the focus of criticism at the expense of recognising the potential of the (perhaps hundreds of) factors that the simulator simulates well. This could lead to undermining its recognition as a valuable educational tool with a clear place in the dental curriculum.

2.5.3 Simulating and Evaluating Realism

Within the existing works, it appears prima facie that a 'selective realism' is considered acceptable and concessions are often made without investigating the implications of their omission or absence. As discussed above, the hardware is far removed from the real clinical environment, is unable to replicate all of the forces involved, in many cases requires the user to adopt a non-realistic posture and hold an unrealistic interaction device. This is barely acknowledged within the literature, however, the realism of the software is considered imperative and any shortcoming deemed unacceptable. But whilst shortcomings are considered unacceptable, many additions have been made as 'features' that introduce differences to what would normally occur in the real world. Equally, some concessions have to be made to recognise the limitations of the simulation environment. Where the line is drawn as to what is a desirable educational addition, what is an acceptable compromise and what the costs and risks associated with these is unknown (Ruthenbeck and Reynolds, 2015). Examples of such features are presented below:

- In Wang et al. (2012) they criticised their simulator because the haptic arm was used to rotate the view relative to the user, which clearly differs from the physical reorientation that an operator would perform to change their view in the real world.

- Debris are created when cutting dental tissue and these particles respond based upon the physics operating upon them as they break away from the tooth. In reality these particles would fall away and be removed with the aid of a dental aspirator. Whilst management of debris is a skill in itself, in VR the removed tissue is not shown and is merely represented with a simple particle simulation to illustrate that debris are being created. (Kim and Park, 2006).

- In some cases, the number of teeth simulated was reduced (e.g. to just 3 teeth in Suebnukarn et al. (2009)). This may be an acceptable compromise as the tooth being operated upon and the immediately adjacent teeth are present, however, the wider oral cavity is not modelled so the skill and challenges of operating in a confined space are not represented.

- In their experimental protocol, Suebnukarn et al. (2009) prevented participants from rotating the view of the tooth. Given the procedure was a crown preparation and the simulator lacked head tracking, this is an unexpected restriction to impose as adjusting the operators viewpoint during the procedure is normally encouraged to ensure correct angulation.

- When visual cues were added to show the reference drilling position and angulation for implant locations these guides led to gradual improvement in performance over time. But after removing the cues, performance reduced (Joseph et al., 2014) leading the authors to rightly question the optimal time to remove the hints.

- To help users appreciate the forces applied by experts, Rhienmora et al. (2010) presented a device which could 'play back' the forces that the expert applied through the haptic arm. By doing so, to accurately follow the path of the 'ghost' handpiece displayed on screen, the student must cancel the system generated force by applying the correct equal and opposite force.

- Similarly, Kolesnikov et al. (2009) reported a device played back the forces applied by an expert during periodontal procedures to assist the learner in appreciating the position and forces used for these procedures. However, this same simulator added a software filter to suppress unintentional hand shaking. Whilst this filter had the desired effect of improving the visual appearance of the simulation environment, it may lead the user to have an exaggerated impression of their hand control.

Even in cases where the simulator is claimed to be a realistic model of the real world, in reality it is a simplification. The complex sounds that a real bur makes whilst cutting through dental tissue is modulated by the speed of rotation, the angle of contact, the power source, the material being cut and the force being applied by the operator. However, the cost of simulating the effect of all these variables would be prohibitive computationally, so a subset is used to provide an approximation of the sound (Wang et al., 2007). Likewise, the haptic feedback of the dental tool is an approximation. Rather than modelling the physical interaction of the bur with the tooth creating the resultant associated debris and gravity, the bur is a simple shape model that is used to collide with and directly delete voxels of the tooth model (Xia et al., 2013). Again, the computational burden of modelling the shear and plastic deformations and tool geometry that would be needed to realistically simulate this interaction would be significant. In both of these cases the simulator designers are trading off the realism, fidelity and accuracy of the simulated world against the performance and responsiveness of the system to maintain a satisfactory visual display refresh rate (Wu et al., 2010). However, when Umemura (2014) investigated the weightings placed on visual information and haptic information when they are in conflict they found the haptic information takes precedence and overrides the visual. This might suggest that reducing the realism of the haptic model to provide computational resource for the visual might not be the optimal use of resources. The validity of these compromises and the extent to which these simplifications impact on performance and perception, particularly of that of experts, is not fully investigated.

The issue of selective realism also extends to issues outside of the simulation itself. One of the stated benefits of VR simulation is that it is a safe environment and obviates the need for personal protective equipment (PPE). As a result just 3 of the studies in this review pictured participants wearing full PPE as would be worn in a real clinical setting. However, the very wearing of PPE may impact upon the perception of the simulator. It has been shown that the wearing of gloves can significantly impact on the perception of touch (Mylon et al., 2016) so their absence may affect the perception of the simulated forces being represented by the simulator. To date the presence or absence of PPE whilst using VR in a dental context has not been discussed in the literature. There are many other aspects of real practice that are often completely absent from the VR simulator such as: the smells created when cutting tooth material, the presence of water spray, safety glasses steaming up due to wearing a face mask or even just the knowledge that the operator is at risk of injury and that there is no ability to undo a mistake.

The above discussion introduces many potential issues. The somewhat arbitrary rules of what must be realistic, what need not be, and where unrealistic interaction modalities can be introduced and their consequent impact has not yet been investigated in a dental context. These semi-natural, moderately realistic, interaction techniques can result in performance levels that are worse than both high-fidelity *and* low-fidelity simulation (McMahan et al., 2010, 2016) so perhaps this requires a choice; to strive for realism, or accept the limitations of current VR simulators and adopt a different focus. The way in which the realism of dental simulation systems has been investigated to date is discussed in the next section.

Evaluating realism

How 'realistic' clinicians and dental students found the simulation was a question asked by 16 of the papers in this review. In most cases, this was part of a wider questionnaire on the participant's simulator experience with a single question of agreement against a Likert scale to indicate the perceived realism. In most cases the questions were variations on 'is it realistic?' such as, for example, "The virtual environment created by the haptic simulator is realistic" (Eve et al., 2014) or "To what extent is the sensation provided by the simulator similar to drilling in a real tooth" (Ben Gal et al., 2011). However, the appropriateness of this assessment method is debatable. *Realism*, or how realistic a simulation is considered to be, is a subjective judgement based on a combination of factors. So a question of how realistic, or life-like, a simulation is considered can be interpreted in many different ways. From a single question it cannot be known if the response is based on the visual appearance, the haptic sensations, the audio, a balance of these factors, or even a subjective representation of the simulation-world inside the user's head (Wages et al., 2004). These aspects of realism can be judged on a continuum (high to low) representing their *fidelity* or the extent to which they match the real-world attribute they represent (Hindmarsh et al., 2014). This ambiguity may have led to inconsistent results in the answering of these questions as, in some cases, participants were asked to evaluate the realism of the simulation when the object being drilled was not a tooth (Ben Gal et al., 2011). Consequently, if a low-fidelity visual representation is paired with high-fidelity haptic sensation it cannot be certain how these factors were subjectively weighted when coming to a judgement of the simulation's realism.

Other authors judged the realism of the simulation based on comments, observation and ad hoc feedback from dental professionals. Here, feedback was more critical and led the authors to believe that they should revise their system and develop more advanced and realistic models (Tse et al., 2010). However, even when authors followed recommendations as to the appropriate level of fidelity for a given educational system, negative feedback about the intentionally removed features tended to lead authors to, again, conclude that the simulation must be more realistic (Heiland et al., 2004). In this particular study it is worth questioning why this feedback was received. The focus of the complaint received by Heiland et al. (2004) was the absence of soft tissues, however, the procedure being simulated did not involve soft tissue, so does this mean a degree of realism is necessary for face validity, or was there a shortcoming in the framing that should have explained that only relevant structures were simulated?

Four papers broke down the questioning about the realism into different categories and produced more thorough explorations of their simulations. Here the visual realism was separated from the feel of the teeth and gingiva (Steinberg et al., 2007; Wang et al., 2015b), or throughly broken down to the feel of individual structures (Wang et al., 2012) and the movements in the oral cavity (Steinberg et al., 2007). However, in all cases, numeric ratings were used and the questions asked were interrelated; meaning the perception of one aspect could impact the score given to another. This may explain some of the low range of responses when the realism is broken down in this way (Wang et al., 2015b).

The issues relating to realism, to an extent, originate from the desire to create an experience that recreates a realistic representation of the oral cavity. Whilst some limited evaluation of dental VR realism has been carried out, the more fundamental question remains of if there is actually a *need* to be realistic. The next section discusses works that did not use exercises based on a simulation of features of the oral cavity.

2.5.4 Alternatives to teeth

Psychomotor tests can reveal the development of dental operative skills missed by traditional assessment methods and performance at these tests improve as operative skills are developed (Shahriari-Rad et al., 2017). However, the value that students get from the use of VR is not equal and varies on the students egocentric visualisation abilities (Kozhevnikov et al., 2013). If the value of a VR experience can be affected by external factors but an abstract psychomotor test correlates with the development of dental operative skills, then perhaps the simulation of teeth is not entirely necessary and alternatives could be investigated to see if they provide a better focus for VR simulation in the dental curriculum.

11 papers used non-realistic oral structures to test of the use of VR in dental education. These papers reported the use of geometric shapes where participants were instructed to remove an indicated area of a block or judge the density of material. Despite the lack of a realistic substrate, tutors recognised that the simulation would be a useful tool for learning perceptual judgement, working with a mirror and planning cavity preparation (Ben Gal et al., 2011) and have wider use for planning implant placement (Cormier et al., 2011).

Non realistic structures were used to familiarise a user with the simulators operation (Eve et al., 2014) and the representation of bone densities (Joseph et al., 2014) prior to caries removal and implant placement exercises respectively. They also feature prominently in the development and evaluation of simulator features where they are used to define the forces for probing for caries (Wang et al., 2016b), material density (Wang et al., 2016a), the value of stereoscopic 3D (de Boer et al., 2016) and haptic feedback (de Boer et al., 2017). In all of these cases, there is an implicit equivalence; that performance and experiences gained on the abstract exercises will be transferrable to exercises involving simulated teeth. This appears to be supported by evidence that abstract exercises are a reliable measure of manual dexterity. Even when these tests find an unsatisfactory distinction between dentists and non-dentists, the suggestion is to revise the pass/fail criteria - not that the test should be replaced with a tooth (Ben-Gal et al., 2013).

The utility of non-realistic shapes was demonstrated in a two-part study where performance on a simulation exercise was used in conjunction with the Perceptual Ability Test (PAT) scores as a predictor for an operative pre-clinical exam (Urbankova and Engebretson, 2011). This showed that when performance on the PAT was combined with performance on a VR exercise it led to a more reliable predictor of future exam performance than either could offer individually. The predictive ability of this test was further improved by the development of a more complex folded torus exercise (Urbankova et al., 2013) which led the authors to suggest that the system could be used to identify which students would benefit from early, pro-active support with their motor skills. It was even suggested the test could be used to screen those students who do not have an innate aptitude for these skills from being admitted to the course. These works suggest that VR exercises need not be realistic to be a useful tool for predicting and measuring student abilities.

An interesting middle ground between realistic oral structures and abstract shapes was presented in the development of 'caries blocks' (Osnes and Keeling, 2017). These cube shaped blocks contain enamel, dentine and pulp layers plus realistic patterns and densities of caries. Despite being an 'unrealistic' shape these exercises were designed to teach the cognitive process of cavity design and strike a balance between enough realism to show face validity and be welcomed by students, but not introduce confounding affects of a realistic tooth interfering with understanding this important concept.

These works suggest that the drive towards realism may not be entirely necessary and non-realistic oral structures can be a useful and effective tool in the VR modality . Their use also removes the confounding effects of realism identified above allowing the user to focus on the underlying educational concepts without the distraction of judging if the simulation is realistic enough. This leads to the next question: if the simulation does not need to be realistic are there benefits if it is deliberately unrealistic?

2.5.5 Does it need to be realistic?

Realistic simulations have been seen to reduce the performance of experts to a level where there was no statistical difference to that of novices (Wang et al., 2015b). This could be explained by the presence of an expertise reversal effect, which can occur as an expert tries to reconcile differences in a new learning experience with their existing knowledge (Kalyuga et al., 2012) but could also be differences in egocentric abilities when using VR (Kozhevnikov et al., 2013) that allow a novice to outperform an expert. If this is the case then the content validity of the 'simulation' starts to be questionable as, all things being equal, a valid *simulation* of a dental operative environment should permit an expert to reliably outperform a novice.

If a realistic simulation can produce these effects and there is evidence of the utility of non-realistic simulation, then perhaps realistic simulations are resulting in missed opportunities that a *deliberately* unrealistic simulation could bring?

As discussed above, effects similar to the uncanny valley can emerge as the visual fidelity increases so, perhaps by deliberately basing an exercise on less lifelike or abstract models some of these issues could be avoided. However, 'artistic representations' could simply be seen as insufficiently realistic (Tse et al., 2010) and users may be surprised at the omission of some aspects and request that they be added (Heiland et al., 2004). The counterpoint to this is that, to provide effective learning environments the simulation should be restricted to relevant structures (Alessi, 1988) so that users can focus on the improvement of individual aspects of performance (Ruthenbeck and Reynolds, 2015). It should be recognised that realism does not happen on its own and that the tutor has a role to play in showing how the simulation applies to real-world practice (Rystedt and Sjöblom, 2012) so, perhaps if a reduced visual fidelity is utilised, it should not be framed as a *simulator* but as a learning tool that simplifies the task environment to allow students to focus on the underlying skills being taught (Reedy, 2015).

One of the reasons that the fidelity of the haptic feedback is regarded as of such importance is that students are expected to be able to learn to judge the subtle variations of force when removing caries (Wang et al., 2016a) and the sensation of cortical breakthrough when preparing the jaw for an implant (Syllebranque and Duriez, 2010). However, exercises can be developed to improve discrimination abilities and provide an environment where optimal strategies can be learned (Jamieson et al., 2015). Perhaps, a non-realistic environment where the sensations are initially presented in an exaggerated form then, as discrimination abilities improve, are restored to their natural values might lead to improved performance or accelerated performance in operative tasks.

A tantalising hint that this may be the case was presented in Suebnukarn et al. (2011) where the effectiveness of VR for training endodontic access preparation was evaluated. Here, two groups were compared after training; one group receiving training using VR and another trained using real teeth on a phantom head. Whilst the authors expressed disappointment that the VR group did not outperform the traditionally trained group in the post-test, they noted that the VR simulator only provided a high speed handpiece and there was a significant reduction in the material removed by the VR group. The authors suggest that this may be due to the simulator group having to be more careful during training, which led to better handpiece control resulting in a more cautious, and consequently, more conservative approach when their skills were applied to real teeth. The significance of this result is somewhat overlooked in the original paper, but it offers a hint that by *modifying reality* in the simulation environment that an improvement might be produced in the performance of the real-world task.

Finally, as discussed above, approximations of the real sounds a dental bur makes as it encounters various tissues within the mouth have been simulated (Wang et al., 2007). However, adjuncts to the audio feedback provided by the simulator could also include 'unrealistic' alarm sounds when the user makes contact with the wrong tooth or if they apply excessive force (Deshpande et al., 2012). Whilst obviously, these auditory cues wouldn't be encountered in reality, augmenting a users interactions with additional feedback introduces new ways of instructing or informing a user as to their performance.

The desire to create high-fidelity realistic simulations is prevalent in the literature, however, these limited (and in some cases accidental) results suggest that there may be an opportunity to develop learning tools which, rather than simply "mimicking the real world" (Escobar-Castillejos et al., 2016) use deliberate and considered unrealistic interactions to enhance or accelerate the acquisition of operative skills. Indeed, the removal of some aspects of the real world can emphasise the features that matter and remove distractions leading to simulation experiences which, despite being of lower fidelity actually have higher authenticity as a learning tool due to the better signal-to-noise ratio (Gilbert, 2016).

2.5.6 Overview

The use of VR for dental education has been driven by a desire to be as realistic as possible. However, the impact of the differences in fidelity to the real environment are unclear. This push towards increased realism may have introduced undesired side effects, which compounded by the limitations of the hardware, has led to uncontrolled and unmeasured variation between the real operative environment and the simulations that attempt to recreate them. Many authors acknowledge that VR is complimentary to traditional simulation but, if this is the case, then it is worth asking if the efforts to digitally recreate an existing training modality are missing the opportunity to create a system that offers *different* benefits to traditional training approaches. Perhaps, as has been observed in the field of aviation simulation, there should be "a shift in focus from the designing of simulation for realism (and hope that learning occurs) to the design of human-centred training systems that support the acquisition of complex skills" (Salas et al., 1998).

2.6 Feedback and Assessment

Immediate and automated feedback is regarded as an important feature to be present in dental simulators (Wang et al., 2016a; Ruthenbeck and Reynolds, 2015; Escobar-Castillejos et al., 2016), with some authors going so far as to suggest that it is an indispensable feature for simulators to be considered acceptable as an educational tool (Wang et al., 2016a). In this review, 38 of the publications reported that their systems provided automated feedback. Various approaches for the provision of feedback to users were evident in the literature, but these could be organised into the following four categories:

- Target based feedback
- Motion and force tracking based assessment
- Time taken
- Clinically relevant feedback

This section addresses each of these feedback mechanisms and presents areas where these systems might benefit from further research.

2.6.1 Target based feedback

Target based feedback was the most common method of providing feedback to users and was provided by 13 of the simulator systems reported. Target based feedback systems present the operator with a 3D region of target material that must be removed from a larger model using a simulated dental handpiece. The system then provides feedback using a combination of:

1. The percentage or mm^3 of the target region material removed
2. The amount of damage caused to the material outside of the target region (expressed as a percentage or mm^3)
3. How much time was spent on the exercise

These values are sometimes combined to show an overall accuracy or performance score (Ben Gal et al., 2011; Yamaguchi et al., 2013).

Whilst it is recognised that this approach has not been validated (Wang et al., 2016a; Ben Gal et al., 2011) it is frequently used as the basis for wider validation of simulation systems in an educational context (See Section 2.7). However, this approach has a number of limitations which have not been fully investigated which may question its value as a basis for assessment.

To achieve a 'perfect' score using this method of assessment, a user must remove all of the target area, without removing any of the surrounding material. Whilst the score attempts to intrinsically motivate the user to aspire to the highest attainable score (Rutledge et al., 2018), this motivation is not directed towards the acquisition of clinically relevant knowledge. The user simply has to remove the target area as carefully as possible; reducing the score to a measure of agreement between the shape cut by the user and the desired pre-programmed target shape. When operating on an

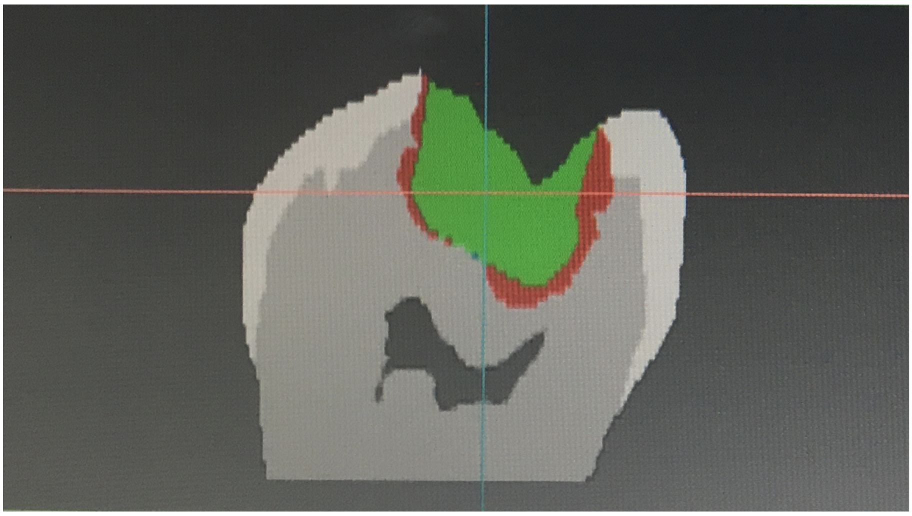


Figure 2.10: A *heavy handed* cavity preparation

abstract shape, this score could be argued to provide a reasonable measure of handpiece control to recreate a desired shape. However, when this same scoring mechanism is applied to a clinical case exercise it can dangerously mislead the operator as to their current abilities and shift the focus to achieving the scores rather than improving their technique (Zigmont et al., 2011).

During caries removal and cavity preparation, the user should be encouraged (or required) to consider essential aspects of sound cavity design such as the smoothness of the margin, degree of undercut, removal of unsupported enamel, flatness or smoothness of the floor etc. (Field, 2016). However, these aspects are not represented in the scoring system, so in order to maximise the score achieved, a user is motivated to instead (perhaps unthinkingly) remove material simply because it is part of the target region without consideration of what this means for the resultant cavity form. This can also incorrectly reward undesirable behaviour such as leaving unsupported enamel where it occurs beyond the target region, because to remove it would further reduce the score. This means that the reasons underlying a a poor performance could be as a result of poor motor skills, a lack of understanding of operative principles, or the user 'gaming' the system, making it difficult to provide appropriate remedial advice.

In the most extreme case it can *completely* mislead a user to their performance. To illustrate this, Figures 2.10 and 2.11 show two attempts at a cavity preparation exercise. In Figure 2.10, indicated in red, it can be seen that the operator was perhaps a little heavy handed and over-prepared, removing slightly too much material beyond the green target area. In Figure 2.11, the operator has slipped and drilled straight down into the pulp chamber, which would be a highly undesirable mistake with serious consequences.

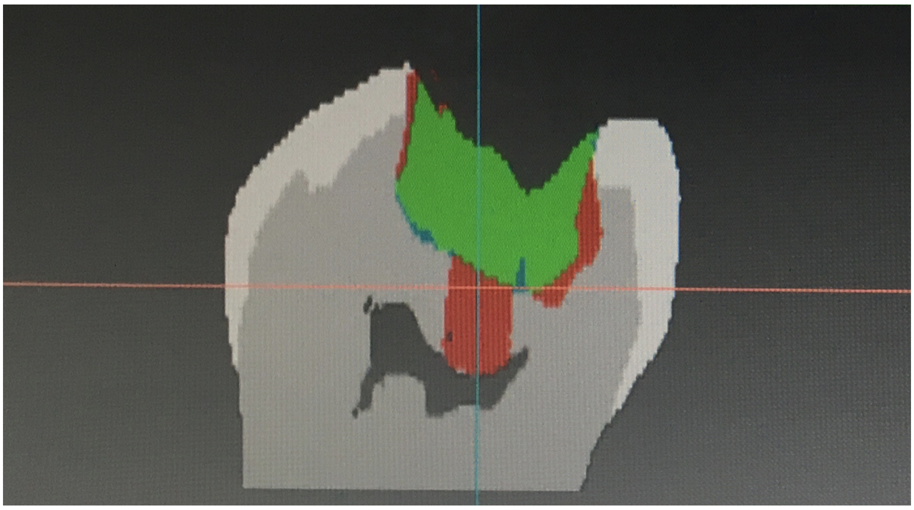


Figure 2.11: A preparation with exposed dental pulp

However, in the case of Figure 2.10 an overall accuracy score of 76.9% was given, but an accuracy score of 85.6% was provided for the attempt shown in Figure 2.11. This metric suggests to the user that their performance was *better* in the attempt with an unacceptable outcome than in one which, whilst in need of improvement, was safely restorable.

The most refined approach to target based training was presented in Osnes and Keeling (2017) which used multiple segmented feedback zones across different areas of a carious lesion. The multiple zones allow for differing expectations of target material removal such as expecting the user to remove all of one region and different threshold percentages of other regions. However, this approach would still 'fail' a user if they removed too much material outside of the target areas without providing them with insight into why this was undesirable. Equally, it would still inappropriately 'reward' failure to correct an uneven floor or margins. This work, however, does represent an important step towards more meaningful feedback and is likely to be the pinnacle of what can be achieved with target-shape agreement based quantitative measures.

2.6.2 Motion and force tracking based assessment

The fully computerised nature of a VR simulator means that the users' motions and the forces applied through the haptic arm can be captured. These recordings can be analysed and compared with the performance of experts attempting the same exercise. The use of motion and force tracking based assessment and feedback was reported by 9 of the studies in this survey.

The most common use of this information is to compare a student's performance to that of an expert. This approach was first presented by (Rhienmora et al., 2009) where the average position, orientation, motion and forces applied by novices were captured and compared to expert users in a crown preparation exercise. This permitted their system to provide additional feedback to the user, beyond the outcome measure provided by the target based approach above, such as "Force minus Z direction should be 2 times higher". Similarly, motion based tracking has been used to train students in endodontic procedures, providing feedback and scoring when their movements differed from the expert performance (Suebnukarn et al., 2010a). After following the guidance provided by the simulator, improvements in simulator and objective tutor measures of their performance could be seen. Finally, this same approach was used to distinguish between novice and expert users (Suebnukarn et al., 2014) when attempting a series of endodontic access cavity preparations. However, there may be limits to this approach. In conducting a detailed analysis of the gestures employed by implantologists, Cormier et al. (2011) found "huge variability" between practitioners, concluding that a typical gesture could not be defined. It is possible that the different procedures under investigation may account for the inability to identify a range of common movements, but this suggests there is varied success in tracking and generalising gestures.

Tracking a users motions has other applications beyond assessment. By taking the same approach, but making it an active part of the exercise, it can add a dynamic explanatory aspect to the experience. Instructors often find that, whilst visual phenomena are easy to describe and explain, tactile sensations are hard to describe verbally (Steinberg et al., 2007). Features can be added to provide active feedback when learning cavity preparation such as arrows to show the direction which force should be applied and varying their colour to show if not enough or too much force is being applied (Lian-Yi and Toshiyuki, 2003). Alternatively, in teaching the forces employed in periodontal procedures the user can actually be physically guided through the procedure (Kolesnikov et al., 2009).

Examples such as these are positive steps to take better advantage of the digital platform and offer new learning opportunities beyond what is already possible with traditional teaching modalities. However, care should be taken to ensure that these features actually enhance skill acquisition. Whilst the force patterns exerted by experts may be measurably different from those exerted by novice users, is this useful information to a novice seeking advice on how to improve? This approach may provide an insightful experience into the forces involved at an abstract level, but the motions and forces applied are the end point of a decision process. Simply training a user to blindly *parrot* the expert users' motions on a specific exercise may not necessarily assist the student in the understanding of the rationale that led to those force patterns being employed so that they can appropriately apply the understanding when presented with a new and unique situation.

2.6.3 Time taken

The time taken to perform an exercise was a recorded metric reported by 25 of the simulators in this review. A computer based simulator makes the capture of start and finish time trivial compared to doing so manually in a traditional phantom head environment. As mentioned above, time elapsed is often combined with other feedback methods to provide a final score. However, when learning or developing a skill, knowing how much time its execution took may not be useful. It is intuitive that in many cases, an expert will be able to perform a task faster than a novice. But, speeding up is a side effect of gaining expertise rather than a prerequisite to become an expert. It is noted that it is not known if faster is actually better, and more accurate movements may take longer (Suebnukarn et al., 2009), so informing a novice that an expert could have done the procedure more quickly without offering any guidance on how to achieve mastery is simply stating a fact that the novice is likely acutely aware of. Additionally, adding this time pressure by making it a factor in the automated assessment may negatively impact the learner's performance and prevent them from focusing on the factors that actually would allow them to improve (Beilock et al., 2004).

2.6.4 Clinically relevant feedback

Finally, 9 papers reported that they provided feedback to users using other clinical measures. These include visual and haptic observations of procedures (e.g. Luciano et al. (2009); Heiland et al. (2004)), comparison of measurements taken in VR with known values (Wang et al., 2012), awareness of wider clinical considerations such as the position of the patient's tongue (Lian-Yi and Toshiyuki, 2003) and realistic clinical measures of performance such as drilling angle, depth, presence of perforations or other iatrogenic damage based on measurements taken in real tissue in a post VR training test (Suebnukarn et al., 2010a, 2011; Ioannou et al., 2015).

Feedback in many of these cases was provided retrospectively outside of the simulation environment in discussions with clinical tutors. Given these differences between the kind of feedback offered by real tutors in a clinical skills lab environment and the nature of the feedback offered automatically, it should be considered what these scores mean and what their role is in the development of operative skills.

2.6.5 The value of a 'score'

Within a dental education setting there is an important underlying aim of teaching reflection and the application of judgement when making decisions (GDC, 2015). As suggested above, the pursuit of a 'high score' on a metric may not actually lead to the desired change and may even inadvertently encourage behaviours that would put a patient at risk if carried out in the real environment. For example, during a cavity preparation a student would be expected to be applying their knowledge and skills in a dynamic environment where they are constantly balancing, sometimes conflicting, factors to arrive at a judgement as to the correct course of action. The assessment mechanisms used in many VR exercises mean that it is questionable if students actually need to understand the science of cavity design at all and can simply develop strategies to "beat" the simulator without understanding the techniques and skill that would be properly associated with that score (Gottlieb et al., 2011).

The presence of a score implicitly suggests that the highest marks are associated with a *correct* way of carrying out a procedure. Whilst there may be readily identifiable *wrong* ways of performing the task or highly undesirable attributes that may be present in its incorrect execution; it may not be as easy to identify a 'right' way of doing something and to encode this into a scoring system. Indeed, what was once accepted as the *right* way of treating caries by removing all traces of infection and 'extension for prevention', gave way to more conservative approaches where the preservation of natural tissue is prioritised (Osborn and Summitt, 1998), so even the definition of what actually is 'right' can shift over time. Equally, in the weighing up of conflicting factors to arrive at an operative decision, there may be cases where there are equally valid arguments to justify alternate paths and different operators may weigh the factors differently and arrive at different conclusions. This doesn't mean that one approach is strictly right and the other is wrong. Consequently, by awarding a score against a predefined notion of 'right', the opportunity for insightful discussions with the learner about the rationale behind their decisions is lost.

If it is arguable about the existence of an absolute 'right' way of carrying out certain operative tasks, then before developing scoring systems to measure aspects of performance it must first be established which aspects of performance actually matter. The more 'abstract' exercises used in Urbankova and Engebretson (2011) make a tacit decision that the precision and the speed of handpiece control are the most important aspects of performance and that an ability to perform a simulated version of real surgical procedure is not the objective of the training. Conversely, the simulator used by Suebnukarn et al. (2014) makes a judgement that *not only* is the performance on a simulated operative exercise important, but also the movements that the handpiece made in order to achieve that performance. In both cases, by providing a score, the designers are 'gamifying' that measure of performance and encouraging users to optimise their behaviour to maximise the score.

The provision of a score on an exercise may be linked to assessment and have implications for a student's progression on the programme. The validity of the scores provided is discussed in detail in Section 2.7 but if the opportunity to use a score in this way is desirable then, to what degree can it be relied upon for this purpose? Differences between the simulation hardware and software and the real clinical environment may mean that there is a degree of construct-under-representation (Messick, 1989) and that performance on the simulator is not representative of ability in the real world. These factors combined may mean that VR simulation and clinical practice are each realities in themselves (Rystedt and Sjöblom, 2012) and that the presence or absence of ability is only applicable in the situation where it was achieved.

2.6.6 The role of simulation

VR simulation in dental education has been utilised in a number of ways. From these applications, 3 potential roles for the use of simulation emerge: as a realistic simulation, a means of evaluating student's skills, or as a teaching tool. The degree to which the accuracy and reliability of automated feedback is an issue, and the level of realism required to achieve a training objective is still an open question (Wang et al., 2016a; Escobar-Castillejos et al., 2016), but it is likely to be governed by the role the simulators will play in the educational setting.

A realistic simulation

As discussed above, the most common use of VR reported in the literature is to attempt to recreate a realistic simulation of an operative procedure to allow a student to undertake additional practice in a safe environment. Phantom heads are currently the most prominent simulation environment for the acquisition of dental skills but these are a partial-task trainer and do not provide automated feedback, instead relying on the expertise of a clinical tutor to compensate for their limitations (Fugill, 2013; Hindmarsh et al., 2014). The extent to which a VR simulator should be required to provide automated feedback depends on if a clinical tutor will also be present in this new environment to compensate for its unique limitations. However, 21 of the papers in this review claimed that a cost saving is possible by dispensing with the tutor and relying on the simulator's feedback alone. If a tutor is not providing guidance, this will mean that the feedback from the simulator is of vital importance. However, at present no simulator reported in the literature is able to completely replace a tutor's presence, as they are unable to monitor or judge a student's posture, use of a safe finger rest, overall conduct or any of the other factors that can materially impact upon performance. By definition, current simulators are only able to identify and grade performance within the terms they are preprogrammed with and as such can't recognise and correct misunderstandings (especially those outside of what the simulator can measure) or remodel the baseline understanding of how a procedure should be carried out (Zigmont et al., 2011).

An evaluation tool

Scores from performances on VR simulation have been shown as a reliable means to identify the students who will need additional support with their fine motor skills later in a course (Urbankova et al., 2013; Urbankova and Engebretson, 2011; Eve et al., 2014) and for discriminating between expert and novice users (Rhienmora et al., 2009; Suebnukarn et al., 2014). As an evaluation tool used for identifying candidates for additional and focussed support, it could be argued that the automated feedback that the simulator offers is not as important. It simply needs to be a reliable discriminator so that additional support can be arranged for any students identified in this way. However, this may not be the most cost-effective use of expensive VR equipment and more conventional testing may be able to judge students with the same level of sensitivity.

An educational tool

A VR simulator can also be an educational tool. Here, the learner is presented with activities and simulations that explain concepts and provide the learner with a safe environment to explore instructional material. As already discussed, these can include experiences through the haptic arm: being directed as to the direction and amount of force to apply (Lian-Yi and Toshiyuki, 2003); experiencing and being able to experiment with the feel of different kinds of bone (Cormier et al., 2011); being guided using the haptic arm to feel the motions and forces used in a procedure (Rhienmora et al., 2010); or a heuristic model for learning the process and motions involved in a tooth extraction (Wang et al., 2015a). In these cases, the simulator is able to tailor the feedback provided to focus on the learning objectives relevant to the concept or procedure being taught without, necessarily, being concerned with creating generalisable measures of performance to be used in a wider context.

It can be seen that the requirement for reliable simulator generated feedback changes depending on the role the simulator is being deployed into. Equally, the same feedback requirement varies in line with the role tutor will play in the simulation context.

2.6.7 The role of the tutor

The lack of availability of tutor supervision is an often cited as an area where simulator generated feedback can be advantageous (e.g. Xia et al. (2013); Ruthenbeck and Reynolds (2015); Roy et al. (2017)), however, as discussed above there are limitations to the feedback that simulators are able to provide so it is unlikely to completely obviate the need for the tutor (Roy et al., 2017). Often, these metrics are ones that can be calculated based on aspects of the performance, but in most cases these scores only indicate the presence of the problem itself. It is an established principle of educational feedback that it should be timely, meaningful and actionable (Larson et al., 2013) to allow the student to recognise where they have made a mistake and what steps they need to take to develop their skills further. There are very few examples of this kind of feedback being generated automatically in a dental context. The simulator reported by Rhienmora et al. (2009) was able to provide actionable feedback in relation to the forces applied (e.g. "Force minus Z direction should be 2 times higher") and this was considered useful by the study participants. However, this is still another quantitative measure of performance that does not explain the issue in the way that a tutor naturally would (e.g. including why more or less force should be used) so may lack some of the authenticity that a real tutor can bring to the simulation scenario (Rystedt and Sjöblom, 2012).

Students who have exposure to VR simulation are recognised as more able than students who have not had the same exposure, however, they are not as able as tutors expect (Gottlieb et al., 2011). This could be as a result of the students not receiving the kind of feedback that they would in a traditional setting so may not have developed their skills in the same way as the equivalent time spent in a clinical skills lab environment. It could also be that the students have become accustomed to the immediate and anonymous feedback from the simulator and find receiving face-to-face feedback from their real tutors intimidating (Gottlieb et al., 2011).

It has been argued that educators should protect the confidentiality of individual performance in simulated context (Zigmont et al., 2011). Whilst this is an understandable desire, as a new pseudo-clinical environment, there is the opportunity to create a "micro-context" (Boud and Walker, 1998) where reflection is encouraged and sharing experiences with others is expected. In this, the tutor is a critical component in the establishment of the norms, and increasing the value that students place in reflecting on that feedback to drive future improvements (Lin et al., 1999). However, unlike in a clinical skills lab, there are two sources of information and feedback; the simulator and the tutor. This may lead to situations where a tutor can't explain or fundamentally disagrees with the score or feedback provided by the simulator. This new tension in the teaching environment introduces an uncertainty as to if the student will put their faith in their tutor's or the simulator's assessment of their work. But perhaps, this tension could present an opportunity to engage in insightful discussion with the student and their tutor will be key to facilitate this opportunity.

Finally, novices often cannot adequately evaluate errors on their own because they lack the ability to detect and apply appropriate corrective actions (Guadagnoli et al., 2002). When coupled with the above limitations in the simulators ability to assess, without a tutor's presence and guidance a learner will have an incomplete picture of the factors that may be resulting in their unsatisfactory performance. Perhaps a blended approach of instant simulator feedback focussed on highlighting areas for discussion, coupled with the insight of a tutor (Roy et al., 2017) will lead to the optimal outcomes.

2.6.8 Overview

There are a number of feedback mechanisms evident in the literature, and each has its own limitations. The desire to reduce or the replace the need for a clinical tutor against this backdrop is perhaps at best, premature. Future work should address the usefulness of this feedback, so that it either more closely matches the feedback a clinical tutor would offer, or alternatively, the focus of the feedback should be shifted so that it can work hand-in-hand with the presence of the tutor. This approach could provide additional insight to the learner to self identify their own areas for development and also provide the tutor with the tools they need to identify focussed areas of support.

2.7 Validity

Attempts to demonstrate the validity of VR for dental education were attempted by 9 authors, plus a further 4 who evaluated simulators without considering their validity explicitly. It is important to validate the use of VR simulation in dentistry both as a tool for learning and in terms of the automated assessment provided by the simulator so that any decisions made on the basis of performance using them can be relied upon as representative of the student's ability.

2.7.1 What is validity?

A test is traditionally considered as valid if measures what it claims to measure (Kelley, 1927) and there are four purposes underpinning testing:

1. to measure an individual's current ability - where the test is a reasonable sample of the scenarios the ability is ordinarily demonstrated in,
2. a measure of how well performance on the test predicts performance on a criterion gathered at a later time e.g. showing correlation with future test scores
3. to estimate an individual's status on a variable outside of what is directly tested or,
4. infer the degree to which an individual possesses some trait or quality

(APA, 1954).

Classic definitions of validity broke the concept into 4 aspects (adapted from APA (1954); Cronbach and Meehl (1955)):

- **Content validity** is established by showing that a test's content is a good sample of all the situations or subject matter of the domain being tested.
- **Predictive validity** is a measure of how well performance on the test predicts performance on a criterion gathered at a later time e.g. showing correlation with future test scores
- **Concurrent validity** is evaluated by showing the degree to which test scores correspond with other measures of performance taken around the same time
- **Construct validity** is an investigation of what constructs can explain variance in a test performance. A construct is some attribute which is assumed to be reflected in performance when no direct criteria is available.

A test may demonstrate concurrent validity but it is not necessarily predictive. However, for purposes of establishing validity, the difference between these two aspects of validity is simply the time at which the corroborating measure was taken. Therefore, these concepts were unified into 'criterion validity' (Cronbach and Meehl, 1955), resulting in the definition of the 3 widely accepted aspects of validity. These aspects were presented as 3 parts of one whole when introduced by the American Psychological Association (APA, 1954), but became to be viewed as 3 separate and different things. This compartmentalisation of 'types' of validity resulted in confusion, leading to oversimplification and a belief that a researcher had 3 'chances' to demonstrate validity. In the face of this confusion it was recognised that these terms had perhaps outlived their usefulness (Guion, 1980).

Over time, scholars in educational psychology refined definitions and measures of validity and unified the definitions under one single form of construct validity. Construct validity does recognise 6 aspects that focus on different issues that may otherwise be overlooked. Underpinning the more modern definition is the idea that the appropriateness and usefulness of a score is inseparable from its intended use and the consequences of any actions arising from its use. These different uses and meanings require different evidential requirements to support the use of the test for any given purpose (Messick, 1995).

2.7.2 Establishing validity in pre-clinical dental simulation

Within the literature concerning VR in dentistry, there have been a number of attempts to demonstrate the validity of the simulators as an educational tool. These attempts have focussed on the classic aspects of face, content and construct validity, but in many cases do not cite an authoritative source for their chosen means of establishing validity. In many cases these papers do not cite where they take their definition of validity from at all, cite a secondary source or one that pre-dates modern concepts of validity.

Two studies discuss the content validity of their simulators defining it in terms of: 1) the ability of the simulator to reproduce the technique used in the operating room so that trainees acquire the correct skills (Ioannou et al., 2015) or, 2) simply stating that experienced instructors demonstrated the equipment, so it is assumed to be realistic enough for teaching and evaluation (Steinberg et al., 2007). A statement of content validity should demonstrate, or explain how the author claims, that the simulator captures a representative subset of the subject matter from which conclusions of performance can be drawn (APA, 1954). However, Ioannou et al. (2015) does not make such a claim, and whilst Steinberg et al. (2007) states the basis is the opinion of expert instructors, it is not framed within a corresponding exploration of the factors from which the sampling was taken. Neither of these studies, fully demonstrate content validity, but could legitimately claim 'face validity' (which has been argued should be added to the 6 aspects of modern validity (Bornstein, 1996)). Whilst face validity is not a demonstration of validity as such, is important that the test looks valid to the participants. A lack of face validity can impact on other aspects of the test's validity (Bornstein, 1996) and impair motivation in test performance (Derous and Born, 2005) so it is an important step towards establishing the wider validity of a test.

Construct validity is discussed by 6 studies (Wierinck et al., 2007; Wang et al., 2012; Ben-Gal et al., 2013; Eve et al., 2014; Suebnukarn et al., 2014; Wang et al., 2015b). In all cases, validity is defined by the ability of the simulator to produce scores which differentiate between novices and experts, with an expectation that experienced dentists would achieve higher scores than novices. These scores are based on: the ability to faithfully recreate a predefined abstract shape (Ben-Gal et al., 2013), a measure of deviation between a reference pocket depth and the value measured by the participant (Wang et al., 2012), measurements of difference between the shape cut and an predefined cavity preparation on a realistic tooth (Wang et al., 2015b; Wierinck et al., 2007; Eve et al., 2014), or the path length of strokes made with the handpiece whilst cutting a crown preparation (Suebnukarn et al., 2014).

However, even when using the classical definition of construct validity, these studies do not fully demonstrate construct validity, but rather provide evidence to support their criterion measure. Therefore these studies are more accurately considering the criterion validity rather than construct validity. The construct under investigation might be *the level of operative ability*, and the score generated by the simulators could be considered as a measure representative of this construct. However, sometimes the criterion is no more valid than the test itself, and claiming that the test measures anything beyond its criterion is speculation (Cronbach and Meehl, 1955). These tests *do* show that qualified dentists tend to score higher than novices, but nothing beyond this appertaining to the level of the operative ability can be inferred. Construct validity can only be demonstrated through a network of associated evidence that combine findings to support the score's meaning (APA, 1954; Cronbach and Meehl, 1955; Messick, 1995). These attempts to establish construct validity in dental education predominantly provide a single measurement which, in isolation, is insufficient to demonstrate the validity of the underlying construct. Only Suebnukarn et al. (2014) used scores from two sources (one external to the simulator) to judge the performance of participants.

It is important to recognise when attempting to establish construct validity, that it is not the measurement of people that is taking place, but the measurement of attributes of people (Guion, 1980). Therefore, it is the ability or quality that underlies the test that is of interest rather than the scores achieved in the criteria (APA, 1954). The percentage agreement between the target shapes or the similarity in movement patterns are not the focus of interest, but what meaning can be attached to those. A measure, whilst convenient or computationally expedient, may not be appropriate for the intended purpose (Guion, 1980). Can the scores from tests of this nature be a sound basis for making judgements about a learner's operative dental ability, and what meaning or consequences are attached to those judgements (Messick, 1995)? The question is not if the test is valid, but if the measurements it provides are a sound basis for making inferences (Cronbach and Meehl, 1955).

More recent definitions of validity treat it as a single unified concept, where the score is inseparable from the meaning placed upon it. If severe consequences are attached to a poor score on a simulator test then the test must be validated to the extent that it can bear the weight of those consequences (Messick, 1989). But this unified definition does not preclude it being useful to differentiate 6 aspects of validity to underline issues and assist in the framing of the meaning, usefulness and appropriateness of a test. These aspects can be challenged using the concepts of 'construct underrepresentation' and tests of 'construct-irrelevance' to probe the validity claimed (Messick, 1995). These 6 aspects will each be taken and discussed in terms of the measures of validity found in the literature so far.

Content Relevance and Representativeness: A key component of content relevance is determining what knowledge or skills will be revealed by the test and how representative of the domain they are. From the tests presented above, what *skill* is being revealed by these scores? They are not based on normal measures of competence that a tutor would look for in the real world, but instead they are derived from how faithfully the participant has followed a predefined shape or has employed stroke patterns similar to those of an expert. This may mean that the skill that is being revealed is not relevant to the underlying construct of a measurement of operative ability. In these cases, care must be taken not to speculate too far as to the meaning of the score without corroborating it with other measures (Cronbach and Meehl, 1955). Only Wang et al. (2012) uses a measure where the ability to make an accurate reading matches the same measure in a real world sense.

It is possible that any measure of ability captured by a virtual reality simulator will introduce some element of construct-irrelevant difficulty. Construct irrelevant difficulty is any aspect of the task that makes it more difficult but is not relevant to the task in hand (Messick, 1995). These could include how comfortable the participant is with computer systems, pre-existing experience with the real situation that impacts on their acceptance of the simulation, or discomfort from the simulator's hardware. If these are occurring, then performance on the task may be impacted and result in a score which is not representative of the user's ability.

Substantive Theories, Process models and Process Engagement: The substantive aspect focusses on establishing if the relevant underlying knowledge is called upon whilst undertaking the test, or conversely, the extent to which a learner can perform well on the test without understanding the relevant theoretical concepts that are being tested. In many cases, the score measuring performance on a VR Simulator is not based on the correct real-world performance of the task, so the knowledge of what correct performance looks like may not be needed. This could be due to construct-underrepresentation of the task (Messick, 1995). As discussed in earlier sections, target based scoring does not rely upon the principles of good cavity design when arriving at a score so a learner can simply remove the indicated region and yet be unable to explain why that region was appropriate for removal. Furthermore, if a learner addresses unsupported enamel beyond the extent of the target, this would further lower their awarded score due to the removal of non-target material even though this is the correct action in the real world scenario. A low score should never result from an assessment under-representing an aspect which if it were included would have allowed the participant to display their true ability (Messick, 1995).

The presence of the target also provides elements of construct-irrelevant easiness (Messick, 1995). In its contrasting colour, the target shows the user what correct performance of the task looks like without requiring them to have the underlying knowledge that would allow them to arrive at a similar shape relying only upon their knowledge. The effect of this was clearly demonstrated in Joseph et al. (2014) where a visual cue was provided to illustrate the correct position and angulation of the handpiece when preparing the jaw for the placement of an implant. User performance initially improved through the cue's presence, but as soon as it was removed there was an immediate degradation in performance, suggesting that users were simply following the guide rather than engaging their knowledge of where the implant should be placed.

Scoring Models as Reflective of Task and Domain Structure: The domain of the task should guide the selection and definition of assessment tasks. Whilst cutting a cavity in a VR tooth may be representative of cutting a real tooth, a student would not normally receive feedback from their tutor in terms of the percentage of material removed correctly or incorrectly. Instead, they would be given guidance in terms of the overall form and the presence or absence of desirable attributes. A score based on percentages does not faithfully mirror aspects of the attribute being measured (Guion, 1980). This may make it difficult for the learner to relate the score to what it means in the real world or the consequences of the equivalent real-world action. Furthermore, this active and constantly updated measure of progress could encourage a trial and error approach where the learner repeatedly consults their score to track their progress. Working towards the pre-determined goal-state without regard to what the real-world consequences of those actions might be and what knowledge should be underpinning them.

Generalisabiliy and the Boundaries of Score Meaning: The scores should be generalisable up to the boundaries of where they can be relied upon. If the scores are based from a non-representative sample of the domain, then the ability to use a score to infer ability beyond the content of what the test explicitly covers is compromised. In most cases above, the attempts at validation have only included the provision of operative treatment to molars. Therefore, it can't be generalised if this VR performance will result in an equivalent performance on other teeth in the mouth. Even within the context of a single tooth, generalisability would require the test to cover a reasonable sample of the cases that may be encountered in practice.

It is also worth considering how the scores change over time and how this impacts upon the generalisability of the criterion (Cronbach and Meehl, 1955). In Suebnukarn et al. (2014) criterion scores achieved by novices quickly converged towards the level of experts. This *could* suggest that the simulator is highly efficient at developing skills, or perhaps more likely, it suggests that the criterion is not sufficiently sensitive to fully represent the underlying ability. It is unlikely that near-expert performance could be achieved within a small number of attempts, and if such improvements can be seen so quickly then this would mean that any sample of performance would quickly become obsolete and not represent the participants current ability.

Convergent and Discriminant Correlations with External Variables: An underlying construct should be validated using measures from multiple sources of evidence. Criterion scores should correlate with other measures of performance including those non-assessment behaviours that would normally be evident in examples of good or bad performance at the assessment task. For example, the presence or absence of a secure finger rest, whilst not part of the assessment, should be detectable in scores provided by the simulator. As noted above, only Suebnukarn et al. (2014) used an external measure to correlate performance on the criterion used. Other claims of validity should likewise seek to cross-reference performance with other measures to more fully demonstrate their results.

In some cases, attempts to validate simulators have not fully controlled for other external variables. For example, in Ioannou et al. (2015) participants were split into study and control groups. Both groups performed a pre-test on an ovine model, then the study group received simulator based training prior to a post-test, whereas the control group moved directly from the pre-test to the post-test. The group who received simulator training outperformed the control group, however it is possible that the simulator group simply benefited from additional understanding of the attributes of a good performance then were then able to recreate them using this knowledge in the post test. To isolate the effect of the simulation experience, the control group should have been given time to study models or watch videos of the correct technique. By doing so, the impact of this external variable would have been controlled and allowed for a more representative comparison of the impact of the simulator.

Consequences of Validity Evidence: The consequential aspect of validity should ensure that the weight of evidence supporting a test's reliability is aligned with the consequences of the decisions informed by performance on that test. If VR simulation is being used to teach operative skills, but significant differences between the simulation and the real technique mean that it can't be assumed that trainees are acquiring the correct skills (Ioannou et al., 2015) then a judgement based on such a system may be seen as an unreliable or an unfair basis to infer real world ability. If this score was then used to prevent a student's progression, this would fall short of the evidential requirements to support the valid use of that score for this purpose.

These modern attributes of validity suggest that there is still some way to go to fully establish if VR is a valid way of judging operative dental performance. These considerations can be combined with the unknown effects of incomplete hardware and software representations upon the development of operative skills. The current evidence base questions if VR is the most appropriate way of measuring operative ability. Could other more cost-effective measures achieve the same end without the initial purchase, ongoing support and maintenance costs that a computerised system necessitates?

2.7.3 The process of validation and accreditation

Much of the discussion above has focussed on the exercises themselves and how their validity can be shown. However, it is also worth considering the source of the claimed validity and the review process that it will have been subjected to. Commercially available simulators are provided with a selection of pre-installed exercises. It is fair to question, from what source, can a purchasing institution take confidence that these exercises and the scoring criteria used to evaluate student performance are accurate and valid? In all the cases above, validation attempts have compared a student's performance with that of an expert on the same task. But, there is a suggestion of inter-subject differences in the perception of simulated dental tissue by experts (Wang et al., 2016a), so it is not unreasonable to suppose that this same difference in perception could lead to different end results. If this leads to experts producing a range of legitimately different results, which one should be selected as the "one true performance" to be immortalised as the target performance for future students to imitate? Is this a decision that should be made by the vendors of these systems?

The accreditation of simulator exercises for dental instruction can be compared to the way that scenarios for pilot training are regulated by external agencies such as the European Aviation Safety Agency or the Federal Aviation Agency. In the aviation industry, each simulator exercise is subject to a detailed and formal validation procedure by a qualified simulation expert to ensure that standards are met and that lessons learned on the simulator translate to the aircraft, and guidance is provided to manufacturers to assist them in meeting those standards (e.g. FAA (1992); EASA (2011)). With the presence of dental regulatory bodies in most countries and moves to harmonise dental education (Field et al., 2017) it can reasonably be foreseen, and perhaps encouraged, that a similar accreditation regime should be introduced to dental simulation.

2.7.4 Validity and the purpose of simulation

This discussion of validation returns to the arguments made by Messick (1989, 1995), that the validity of a test cannot be detached from its intended purpose. Returning to the 3 roles in the previous section, it can be seen that the validation requirements for each of the roles will be quite different. As a *realistic simulation*, the validity must be compared to the equivalent real procedure and demonstrate content and substantive aspects of validity. As Ioannou et al. (2015) recommended, the motor skills and cognitive processes developed on the simulator must be transferable to the real procedure. This should also extend to any consequential aspects for patients receiving treatment after a learner's simulation experience. As an *evaluation tool*, the requirements of validity shift to the demonstration of the generalisability of the findings and the consequential aspects for the students being assessed so that the evaluation of the student's ability is fair and that the consequences of a poor score are proportionate to the reliability of the measure. Finally, as an *educational tool*, the simulator must demonstrate its validity with a focus on the structural and substantive aspects to show that the student is developing appropriate knowledge, skills and attitudes.

This evidential weighting based on the purpose of the test can be seen in aviation simulation. As discussed previously, flight simulators are divided into different categories dependant on their intended use. These different uses have different evidential requirements to demonstrate their validity. If the purpose of a simulator device is to provide a full and realistic representation of the full flight experience, authentic enough to count towards flight hours (CAA, 2011), then the requirements are much more rigorous than if the purpose is to simply provide basic familiarisation (EASA, 2011).

2.7.5 Overview

The validation of VR simulation must be framed within the context of its intended purpose with reference to modern concepts of validity. Only through establishing the intended place for VR in the dental education curriculum can it be appropriately validated for that use. These activities will inform the requirements of the scoring and feedback mechanisms to meet that purpose and in turn reveal the hardware requirements necessary to provide them. In time, this should lead to the establishment of minimum standards for simulators used in dental education with appropriate oversight from a regulatory body.

2.8 Gaps identified in the literature

This comprehensive review of the use of VR in pre-clinical dental education has revealed a large number of issues spanning numerous domains that require further investigation. These can be summarised as:

- There are no standards that a VR dental simulator must comply to. This means there is no guidance nor any compliance requirements for a VR dental simulator or its exercises to meet before they are considered appropriate to be used in the instruction of operative skills.
- The absence of standards prevents the reliable transferability of pedagogic insight developed on one simulator being generalised to other devices. This means it is difficult to draw general conclusions as to the effectiveness of VR in dental education from the present literature.
- The impact of the fidelity of the simulator in terms of its visual appearance, the perception of 3D and the ability to recreate operative forces is unknown.
- The hardware, ergonomics, use of PPE and affordances of the simulator differ significantly from the real operative environment and the ways in which this impacts upon transfer to a real clinical environment is unknown.
- VR simulation has been demonstrated as filling a number of roles. Whilst these roles are not mutually exclusive, clearly identifying where VR can be most effective would allow a better evaluation of its appropriateness and validity for that purpose which may inform what the necessary functionality should be.
- Research to date has focussed primarily on the creation of realistic simulators, placing them in competition with phantom head devices. The framing of VR as an adjunct teaching tool with a focus on developing complementary skills is underreported.

- Attempts to validate simulation and its automated scoring mechanisms have been based on simplified or isolated notions of validity and do not demonstrate validity through investigation of the unified construct covering the content, criteria and consequences of their use.

- The educational impact of features such as the ability to undo, reset and view live cross-sectional views and feedback should be investigated to ensure these are utilised appropriately and are not harming the ability of a student to deal with unexpected events.

- The interaction between clinical tutor and simulator-generated feedback should be further explored. A skilled and enthusiastic teacher will generally be able to assist a learner's development regardless of the modality, so exploring how the simulator-generated feedback can enhance teaching and assist learner comprehension will help to ensure the maximum value is gained from this new environment. Conversely, it is conceivable that tensions could be created where a tutor's opinion of a learner's performance differs from that suggested by the simulator-generated feedback. Strategies for dealing with such disagreements or a clearer appreciation of the limitations of the automated assessment should be developed.

- As a new teaching tool there has been insufficient time for the longitudinal impact of these devices to be properly assessed. Furthermore, the novelty of a new environment itself may also be influencing present results. As the use of these devices becomes more established it is vital that longer term studies are conducted to explore their contribution to learner achievement.

- VR in dental education draws upon a number of areas of research including Dentistry, Computing and Education. If these aspects are not given equal consideration it can confound findings leading to erroneous conclusions. For example, if scores and feedback are provided by a simulator that omits subtleties of the operative environment this can mislead participants as to their true ability in the target context. Equally, if the hardware configuration of a simulator has ergonomic differences to the operative environment it could lead to the development of skills that differ from their real-world equivalent and mean that the skill has to be re-learned (and the simulation experience un-learned) when transferring the skill to other contexts.

- Finally, many existing studies are confounded by their choice of measurement instrument. Simulation in a dental context is intended to develop both motor skills and the acquisition of the cognitive knowledge that underpins operative performance. Where a measurement of a simulator's effectiveness is exclusively based on a learner's ability to perform a procedure using a VR simulator it could under-represent the learning that has taken place and not be truly reflective of the simulator's effectiveness. A learner may have acquired the cognitive knowledge associated with the procedure but have been unable to demonstrate it in this context. Given the differences between the simulation and other modalities this may not preclude their ability to successfully demonstrate the skill elsewhere.

Chapter 3

Measuring, Developing and Transferring Learning

Having reviewed the literature concerning the applications of VR in dental education, it is evident that in order to explore their most beneficial use case, wider literature must be consulted.

This Chapter will explore:

- How learning is measured, attributes of high-quality measurements and how these relate to learning and teaching of dental operative skills
- How skills developed in one context are transferred to another and theories that influence the effectiveness of this process
- Broader considerations such as the impact of cognitive load and how practice sessions are structured

Throughout these sections, insight is related to the VR context in order to inform the refined research question presented in Chapter 4.

3.1 Measuring Learning

In order to compare between different educational interventions (and no intervention at all) it is important that measurements against a scale can be taken. From this basis, the concept of 'Learning Gain' has emerged. However, across the literature this term has acquired two major definitions: firstly, as a subject-specific measure of knowledge taught within a course, and secondly, as a measure of a wide range of non subject-specific personal attributes, abilities and skills that are developed as a result of a course of study (Vermunt et al., 2018). However, within these different uses of the term both broadly agree that learning gain is a quantifiable measurement of a change in (and retention of) newly acquired skills, knowledge or attributes between two points in time (McGrath et al., 2015; Pickering, 2017).

The broadest of these definitions suggests that over a programme of higher education students learn 3 things (McGrath et al., 2015):

- The *content knowledge* relevant to the field of study, which makes up a significant part of the programme;
- *Skills and competence*, that is, the ability to apply the practical skills and cognitive knowledge to complete tasks or solve problems related to their field;
- Broader *personal development* such as self reliance and respect for diversity by mixing within a cohort drawn from different backgrounds.

This broad definition has led to significant interest and investment into the approach by government agencies who see it as a way to justify and direct investment in the Higher Education sector and measure its value for money, increase accountability and allow students to compare outcomes between programmes offered by different institutions (McGrath et al., 2015). This work has led to recommendations that programmes of study should clearly signal the core concepts and how they will be learnt, embed a measurable change between before and after the programme of study and use authentic assessment methods that measure what the students *actually* need to know (Evans et al., 2018).

3.1.1 Learning Gain

There are a number of ways in which learning gain can be measured, which were summarised by McGrath et al. (2015):

- *Grades* such as marks and scores from testing. These clearly rely upon them being comparable between two timepoints for example by assessing the same topic.
- *Surveys and questionnaires* where students are asked about their perceived learning gain during the programme.
- *Standardised Tests* provided externally. In a dental context this might include the University Clinical Aptitude Test (UCAT).

- *Other methods* such as allowing students to reflect upon their learning, discussions with other students and tutors or personal development portfolios.

Embedding one or more of these instruments from the outset can allow the effectiveness of the curriculum or an intervention to be measured and improved. However, there are difficulties in showing a causal link between an intervention and any improvement in learning due to the confounding factors of retention, decay, ongoing training and normal maturation (Colt et al., 2011).

Retention of knowledge and decay between time points are a clear concern for an accurate measurement of learning gain but they are perhaps the most readily controllable. Decay can be controlled through careful timing of subsequent measurements and if one approach results in superior retention, this can be revealed through testing. However, testing and examinations only tell us about a student's knowledge at a single point in time and in developing effective teaching interventions it is of more interest to the instructor what happens during the transitions between 'states' (Meltzer, 2002). Programmes of study do not exist in a vacuum; content and skills overlap between modules and it can't be said with certainty that, for example, a student's improvement in their patient management skills is caused by the increased exposure to the clinical environment and guidance of their tutors or by some unrelated factor such as starting a part time job that has allowed them to develop their interpersonal skills (example adapted from Suskie (2004)).

Clearly, the effects of ongoing training and normal maturation are much harder to isolate but even the measurements taken under test conditions are problematic. Performance strongly correlates with factors outside of the instructor's control such as pre-instruction knowledge (where students tend to cluster based on their starting point (Ben-Gal et al., 2017)), time spent preparing and other confounding affects such as exam nerves (Meltzer, 2002). To effectively measure learning, the measurement must have maximum dependence on the instruction provided and minimum correlation with pre-instruction knowledge.

Even taking these factors into account, performance is not the same as learning. It can't be reliably inferred that person A is better than person B at bowling if A got a strike and B got a gutter from a single performance; the differences must be stable and repeatable. We are interested in measuring the underlying capability or capacity for performance, not the result of an individual performance (Schmidt and Lee, 2014, p. 178). Learning is the only factor that *consistently* affects the difference between trials such that insight gained from an earlier performance influences the next one (Chambers, 2012). When comparing performances we can't always infer that learning has or has not happened because the results of that performance can be impacted upon by a wide range of factors such as personal standards, importance placed on the performance or even 'gamifying' the score so that performance is shifted towards an artificial optimal at the expense of un-measured attributes (Chambers, 2012). For example, if time-taken is a criteria, a score could be artificially boosted by 'gaining' speed at the expense of quality.

When testing a psychomotor skill to discover if learning has occurred, there are further questions relating to the environment that the test is conducted in (Schmidt and Lee, 2014, p. 176). Should it be done in a realistic and stressful circumstance similar to the situation in which it will be used, or in a more calm environment as might have been used during training? At what point should the testing be done - at the end of the course when fatigue might be a factor or some time later when the skill has decayed? Should an initial measurement be incorporated and compared to a measurements taken throughout the course, and if so, how many measurements should be taken?

Taking multiple measurements longitudinally (especially as found in pre-test and post-test designs), whilst highly desirable, can introduce 'floor and ceiling' effects (Meltzer, 2002). Pre-test and post test studies measure if participants have improved or regressed over the course of an intervention and then attribute the results to the intervention itself (USDoE, 2003). However, for any given test score, it is impossible to achieve less than 0% or more than 100%. Therefore, performance in the pre-test can limit the measurement of the learning gain across the intervention. For example, a student who scores 50% in the pre-test can see up to a 50% improvement over the intervention whereas a student who scores 85% can only see a maximum of a 15% gain. If a later test shows that they both improved their performance by the same percentage, did one learn more than the other? Also, if the test were harder it is possible that a different outcome would have been found (Meltzer, 2002).

The ceiling effect often leads to a strong negative correlation between absolute gain scores (calculated as post-test score minus pre-test score) and a student's pre-test score because, all things being equal, a higher pre-test score leads to a smaller absolute gain (Meltzer, 2002). A valuable tool to overcome these floor and ceiling effects is the concept of *normalised learning gain*. This approach was introduced in Hake (1998) and has been used to good effect when measuring learning gain in a number of contexts including medical (Colt et al., 2011). This approach, as the name suggests, *normalises* the test scores to give a score that is relative to the maximum gain a student can make based on their initial score. This permits the presence of a wide range of pre-test scores to be present and then after receiving an identical intervention, the normalised gain should not differ significantly.

However, in the case of dental operative skills even a normalised gain score could be problematic. It is often difficult to draw too many conclusions from the pre-intervention measurement because many students enter the programme without any prior knowledge so the floor-score itself might not actually be useful (Suskie, 2004), furthermore, care must be taken to balance the number of testing points because it has been shown that increasing the amount of testing itself can impact positively upon performance in a dental context (Sennhenn-Kirchner et al., 2018).

3.1.2 Desirable attributes of learning measurements

From the above discussion, it is evident that any measurement between two points in time will have issues that are tightly coupled to the specific context in which the measure is being deployed. So, when selecting an appropriate measurements for a given intervention, McGrath et al. (2015) suggests that a robust measurement of learning gain should exhibit the following the attributes:

- Longitudinal or cross-sectional designs are preferable because they provide a broader base of evidence for measurements. These also permit a measurement of the permanence, or retention, of the acquired skill or knowledge.
- Representative measurements using a broad base of participants are important to prevent exceptional cases from skewing the data from the real norm.
- Validity is evident in all aspects of a robust measurement so is a key consideration.
- Comparability, so that they can be used for inter-institution comparison.

The robustness of measurements of learning gain are increased when multiple measurements are taken, which agrees with the principles of construct validity (Messick, 1995), but combining measurements can introduce an element of subjectivity in how they are combined (McGrath et al., 2015). however, the potential gains, richness of data and cross-verifiability possible from a broad base covering different aspects of the intervention would arguably outweigh a risk of subjectivity.

The context in which measurements of learning are used also needs to be considered, so the next section will discuss how dentistry is taught and the regulatory framework it operates within.

3.1.3 Learning and teaching dental operative skills

Within the United Kingdom, the provision of accredited dental degrees is regulated by The General Dental Council (GDC). The GDC define a series of learning outcomes that a graduating student must be able to demonstrate in order to be added to the register of dental professionals. It is up to an individual institution to demonstrate how their students have satisfied those outcomes, but the GDC provide oversight and monitoring of the programmes to ensure that expected standards are met (GDC, 2015). Successful completion of the programme is an assertion from the educational provider that the registrant meets the criteria for a 'safe beginner' and programme inspections will attempt to corroborate these conclusions (GDC, 2015).

The learning outcomes set by the GDC cover the three domains of knowledge identified by Bloom and Krathwohl (1956): the affective domain, cognitive domain and psychomotor domain. The affective domain includes empathy, behaviour management and aspects related to interpersonal skills; the cognitive domain includes the critical thinking, technical and procedural knowledge aspects; and, the psychomotor domain is skills directly related to being able to carry out the procedures and treatments. In order to graduate, a dental student must integrate and successfully apply skills across these domains in the clinical environment (Segura et al., 2018).

It is assumed that dental operative motor learning can be measured on a curve from a beginner to proficiency and its acquisition is based on the premise that these skills are developed through practice and training (Ben-Gal et al., 2017). The journey along this curve begins by undertaking a basic clinical skills course that provides a programme of foundational operative procedures in a simulated environment supported by tutor feedback, instructional videos and printed material (Field et al., 2020). When a student has demonstrated a safe standard of proficiency in the simulated environment, often via a gateway examination, they progress to the provision of these treatments to real patients (Segura et al., 2018), refining their skills through repeated performance across a number of patient cases. When the student is ready to learn additional, more advanced, procedures they return to the simulation environment and the cycle repeats itself.

Whilst delivering the treatments to patients in the clinical environment, most dental schools require that students repeatedly demonstrate each procedure to an expected standard a number of times through the setting of clinical targets (Chambers, 2012). Individual performances of a treatment are awarded a grade by a clinical tutor based on their evaluation of the student's performance on that day. If the grade received is of a high enough standard, it counts as one towards their clinical targets and when they have acquired the number required they are said to have met that target. However, the rationale for the setting of a clinical target at a specific number (as opposed to one more or fewer) to demonstrate competence is regarded as somewhat of a "traditional mystery" (Chambers, 2012).

The challenge of measuring dental operative performance

The premise that repeated practice of a given procedure is based on is that over time there should be a measurable change in knowledge and technical skill (Colt et al., 2011). However, the 'law of practice' (Snoddy, 1926) states that there are diminishing returns over time resulting from practice. Large gains can be seen early on but only minor improvements will be seen later. Indeed, once dental students are providing operative treatments, they require an average of 125 extra procedures worth of experience to boost performance one point on the grade scale (Chambers, 2012). Additionally, improvement in ability does not come solely from repeated practice at a procedure. Whilst measurable increases in performance can be seen early in the learning curve, they will eventually reach a plateau (Ben-Gal et al., 2017) and by the time a dental student is considered to be at a level of proficiency to be safely working with patients they are already in the maintenance phase of performance and there is no evidence that repeated practice beyond this point further improves performance, nor is it predictive of test performance and the learning curve is, essentially, flat (Chambers, 2012). It also neglects to incorporate the shift in learning goals at isolated procedures from beginner to a more competent performance which requires the integration and management of procedural skills with wider needs including a focus on the patient more holistically (Chambers, 2012). This means that measurement systems that fail to take this shift into account are not truly reflective of the learning that is taking place. Even comparing students with the same volume of experience in an undergraduate setting must only be done with caution because it also fails to capture the context or difficulty where that experience was gained such that any such comparison can be considered valid (Dawson et al., 2021). A much stronger correlation is found between a student's self-assessment ability and their preclinical performance (Lee et al., 2017). Given that improvement in ability draws from both the motor learning and the gain in theoretical knowledge (Ben-Gal et al., 2017) plus the need for students to graduate as self aware, self regulating, reflective, practitioners (GDC, 2015; Lee et al., 2017), an ability to recognise their own ability and what needs correcting within it is a natural precursor to better performance. However, despite this and the prevalence of a great many objective criteria to judge performance in dentistry, it is a skill that students are often poor at (Lee et al., 2017).

It emerges from the above that there are challenges to measuring learning of competence in dental operative procedures. There are two environments in which competence is expected to be developed and demonstrated:

- Firstly, in the simulation environment where cognitive and psychomotor skills development must be demonstrated to satisfy a gateway examiner that the learner is safe to progress.
- Secondly, in the clinical environment with real patients where the affective domain is introduced and the learner must integrate knowledge and procedural skills whilst accommodating the stresses of patient management in a real clinical environment (Segura et al., 2018).

The premise that simulation based training is based upon, and the very notion of practice itself, is that permanent skills can be developed in the training environment to allow the individual to solve real problems of that kind when encountered in the target context (Schmidt and Lee, 2014, p. 199). Virtual reality was introduced in to dental education, in part, as a tool to provide students with opportunities for additional repeated practice with no incremental costs (Xia et al., 2013) permitting more experience to be gained during the simulation phase of training. But, just as with traditional simulation, skills developed in VR must assist with preparing the student for the real clinical environment - the skills must *transfer*.

3.1.4 Transfer of skills

Measuring learning gain on some scale is clearly valuable, however, ultimately the purpose of the dental training environment is to prepare a student to provide their first treatments to patients; not to simply be able to perform well in the training tasks. Transfer is one of the central issues in learning (Holding, 1989, p. 284) and is the study of how knowledge acquired in one situation transfers (or fails to transfer) to its use in another (Singley and Anderson, 1989, p. 2). The use of simulation for training permits a measure of the extent that a skill learned with a simulation device can be expected to allow the learner to perform the trained skill with the real instrument (Hammerton, 1967). Practice with a simulator can have no effect or can be considered to result in positive or negative transfer. Positive transfer improves performance in the target environment, negative transfer degrades it. But the real value of the skill acquisition activities is best measured in the context the learner will ultimately need to demonstrate it (Salmoni, 1989, p. 218). Therefore, any simulation experience that results in an improved performance in a clinical environment can be regarded as an example of positive transfer, but any shortcoming of the simulation experience that results in confusion or uncertainty when relating the simulation to the real-world would be regarded as negative transfer.

There are a great many theories of how to measure that transfer has taken place, but they can be broadly categorised in to 'savings' and 'first shot' measures (Hammerton, 1967). A savings measure tests how much time (or practice) is saved using the real equipment by practicing on the simulator. A first-shot measure asks how will the learner respond the first time they encounter the situation after benefitting from the simulation training. Tests of transfer allow the results of an intervention to be expressed in terms of an estimate of performance on the criterion task and can be expressed as a percentage transfer (e.g. difference between groups) or as a score expressing how much 'practice' has been saved due to the intervention (Schmidt and Lee, 2014, p. 191).

As mentioned above, when investigating psychomotor skills leaning, the purpose of the learning is to be able to apply it in the real world. So the design of experiments must take in to account the effects of manipulating variables to test if learning has occurred. These results can often be influenced by the participant's knowledge of the expected result and not capture the desired measurement of the more permanent learning or their ability to use that knowledge or skill in its intended context. Therefore, transfer designs to measure the effects of learning new skills are considered critical for the study of motor learning (Schmidt, 1975; Salmoni et al., 1984).

3.1.5 Overview

The measurement of learning is complex, and even more so when attempting to measure the performance of a task in a domain which encompasses theoretical knowledge, fine motor control and inter-personal elements. But this very complexity makes it most suited to being measured through transfer as multiple aspects can be investigated and measured. The use of virtual reality simulation as a vehicle for enhancing transfer may prove more useful in some areas than others. How transfer can be facilitated and how psychomotor skills are actually learned will be discussed in the next section.

3.2 The Development and Transfer of Psychomotor Skills

One of the most frequently cited definitions of *skill* was provided by Guthrie (1952), who defined it as "the ability to bring about some end result with maximum certainty and minimum outlay of energy, or of time and energy". This sentence concisely embodies a number of important points. To demonstrate skill, there must be an intention for the result and it must be reliably reproducible. A novice and a professional both playing a game of 10 pin bowling may both score a strike (which both intended) but, for the novice there is likely to have been an element of luck and it is unlikely that this could be repeated reliably and with any degree of certainty. The minimum outlay of energy is also an important inclusion. A professional runner will be able to maintain a greater speed with a reduced level of energy expenditure than a novice through superior technique. This consideration also applies to the mental energy expended during a task because a skilled performer will have additional cognitive resources available for planning and strategy (Schmidt and Lee, 2014, p. 6).

There are 3 elements that are critical to possessing any skill (Schmidt and Lee, 2014, p. 8):

- The ability to perceive the relevant environmental features
- Making a decision on *what* to do, *when* to do it and *where* to do it in order to achieve a goal
- Producing the organised muscular activity needed to generate the movements that will achieve that goal

Of these, the most important feature of a skilled performance is deciding what to do (and equally, what not to do) in a given situation. Possessing the most precise of motor skills is of little value if they are employed incorrectly and an undesirable result is produced through poor decisions stemming from a lack of knowledge. Schmidt and Lee

(2014) present a sequential model for information processing leading to a skilful response. It is elaborated upon over the course of the book, but the basic model (Schmidt and Lee, 2014, p. 22) presents, between an input and an output, 3 processing stages that neatly map on to the critical elements of skill above:

- Stimulus Identification
- Response selection
- Movement programming

In the stimulus identification stage a performer must first recognise that a stimulus is present and then assess what it is by using the available channels of perception such as sight, sound, smell, taste etc. Next, the performer then must decide upon and select an appropriate response to the stimulus. Finally, they must organise their limbs to produce the desired movement.

As the number of available stimulus-responses increases in a complex environment presenting many available options, the processing time required to select the correct response also increases. Additionally, the capacity for information-processing is finite and information can be missed when it is overloaded (Schmidt and Lee, 2014, p. 40). Furthermore, focus on the incoming stimuli tends to be serial; focussing on one thing and then another so it is a crucial skill to know where, and on what, to direct this finite attention.

At the core of focussing attention, is the ability to perceive the *relevant* environmental features in order to gain insight into the nature of the problem being faced (McCloy, 1968). Success in the execution of a skill is often determined by how well the performer detects, perceives and uses the relevant sensory information (Schmidt and Lee, 2014, p. 64). However, when training in the real world, extraneous stimuli compound the difficulty of identifying and associating the correct response because the performer must decide what is irrelevant and can be disregarded (Hammerton, 1967). But a virtual reality simulation could reduce these distractions and capture the *essence* of the stimulus (Singley and Anderson, 1989, p. 118) aiding the learner in identifying what stimulus should be driving the decision making process such that when the same stimulus is encountered in a different context they will select the appropriate response. Therefore, the test of the effectiveness of the simulation environment is how well it develops the means for a learner to generalise and apply the knowledge gained in training and select appropriate responses in the real situation (Kozlowski and DeShon, 2020).

3.2.1 The acquisition of psychomotor skills

The purpose of *learning* a skill is to do something different which leads to a (hopefully improved) change in movement patterns. When learning a psychomotor skill, the most important factor is practice, and providing increased opportunities for this was a key driver for the introduction of VR to dental education (Xia et al., 2013). More practice leads to more learning and to the development of 'motor programmes' (Schmidt, 1975). Motor programmes are a series of steps, or subroutines, that describe a motion and are stored in long term memory for recall and execution when the skill is needed. However, it is important to recognise that not all practice is equal and simply increasing practice time should not be the focus. It was noted earlier that skills can plateau, but effective practice is more than simple repetition, it is solving a *new* problem using experience gained from similar situations in the past (Schmidt and Lee, 2014, p. 199). However, there is often a conflict during practice sessions between learning as much as possible and performing as well as possible. Often feedback is based on mistakes so by definition is not given out equally. Unless the information is covered elsewhere, if a student's performances never produce a particular error then they may never be instructed that it is something to avoid. Therefore, "doing your best" may be undesirable from the standpoint of maximising learning as it prevents experimentation between attempts and precludes the feedback that would be associated with it (Schmidt and Lee, 2014, p. 200). This has led to the suggestion that there should be distinction between practice and test sessions (Schmidt and Lee, 2014, p. 200). In practice sessions, the learner should simply avoid repeating what they did earlier and try different styles of movement. Instruction and feedback can be provided to avoid inappropriate movements but the learner should know that the focus is not on the quality of the output. Then, test sessions measure the progress of the skill that was developed in the practice sessions.

The focus on achieving goals may also impede the desired learning. If a performer's entire cognitive capacity is devoted to the achievement of a goal then there is no capacity left for appreciating the steps that they have taken to achieve that specific performance so that they can be encoded in long term memory (Sweller, 1988). As discussed in Section 2.6, the use of target shapes and percentage accuracy scores in VR exercises may be undesirable and could be promoting goals that do not lead to learning. Such tasks focus the learner's attention on the goal of removing 100% of the target area whilst minimising the damage to the surrounding material. This is likely encouraging the learner to deploy a means-ends strategy (Sweller, 1988) to achieve the goal (by repeatedly comparing the current state to the desired state and then taking steps to reduce the differences). If the learner's entire cognitive capacity is focussing on the removal of the identified target material to achieve a high score, they are unlikely to be developing desired cognitive skill of recognising the desirable attributes that the target shape was supposedly demonstrating.

Learning is an important goal of practice and it is important that the knowledge and learning acquired in practice can be used in other situations. Knowledge can be classified as declarative or procedural (Anderson, 1981). Declarative knowledge is factual and easily verbalisable: 'knowledge that' for example a feature is undesirable or that there are 26 letters in the English alphabet. Whereas, procedural knowledge is a 'knows how' skill or activity that can be demonstrated but is often harder to communicate. For example, Polanyi (1958, p. 49) states from his discussions with physicists, engineers and bicycle manufacturers that the principles by which a cyclist can balance on a bicycle were generally known but, whilst a rule can be derived, knowing it does not automatically enable one to not fall off. In a dental context, a student may know that a degree of retentive undercut is an important feature of the cavity design for an amalgam restoration (declarative knowledge) but not have an idea how to successfully apply that principle when preparing a cavity using a dental handpiece (procedural knowledge). However, whilst the skill is demonstrated procedurally, the declarative knowledge is the reference that the learner uses to judge their own performance, such that it forms the basis of transfer for different uses of that same knowledge in different contexts (Singley and Anderson, 1989). Therefore, refining and perfecting the declarative representations is key to the transfer of the knowledge to its intended context and facilitating this should form part of practice sessions.

However, learned skills are not retained indefinitely and a failure to practice them will lead to a decrease in performance. Declarative knowledge or tasks with a larger 'cognitive' component are forgotten more quickly than procedural knowledge or tasks of a more 'continuous' nature. The classic examples of these being the differences between riding a bike versus remembering a phone number (Singley and Anderson, 1989, p. 198). In psychomotor skills development it is important that training is able to bridge this gap and enable the student to apply the information received in declarative form to be applied and demonstrated procedurally to assist in its retention. This is an area that VR simulation is well placed to assist, as taught content can be provided and quickly followed using specific exercises to allow the immediate procedural integration of the knowledge into a relevant task, aiding retention.

Practice also has a number of other benefits in addition to the obvious aim of improving performance (Schmidt and Lee, 2014, p. 200):

- It increases perceptual skills so that a learner can detect information about the environment and feed this in to their decision making.
- It leads to a reduction in capacity demands; well developed skills become more automated so don't require the same attention to be paid. This frees up capacity for strategy and planning.
- With the benefit of practice, error detection is improved and a learner who can detect their own errors will correct them without feedback which leads to an increase in skill.

- Finally, ‘coarser grain’ motor programs are developed.

For example when learning and practicing how to change gear in a car, the learner driver has to consciously remember to 1) Lift the accelerator 2) Depress the clutch 3) Move the gear lever to neutral 4) Move the gear lever to the selected gear 5) Lift up the clutch 6) Depress the accelerator. However, with experience these movements become unified in to a “changing gear” motor program that the learner can give effect to without consciously monitoring the individual steps. This means that the resources that were previously consumed by recalling and sequencing the steps to change down a gear when approaching a junction can be re-directed towards oncoming traffic to enable advance-decision making which enables a smoother turn.

Finally, the role of instruction in the acquisition of a skill is of vital importance. Direct instruction in the form of spoken, written, pictorial, or demonstrations provides information on the first aspects of performance, including, for example, how to position oneself, how to hold the apparatus, where to direct attention, or what to expect (Schmidt and Lee, 2014, p. 231). These early instructions are crucial in improving skill levels in early performance, however, they are often provided in complex environments closely resembling the environment in which they will be deployed. Yet, as argued by Anderson et al. (1996) this does not necessarily need to be the case; an accountant must develop the interpersonal skills to deal with their clients, but they do not need to learn to do this at the same time as learning the laws and regulations surrounding taxation or how to use accounting software. Recognising this may be the case and making appropriate separation of tasks often falls on the tutor so that the learner’s attention can be directed and appropriate instruction provided.

3.2.2 The stages of learning

The above discusses the attributes present in skilled performance and the process leading to that skilled performance. However, there are stages and different levels that the learner goes through as their skills develop. These are perhaps not fixed states as such but are valuable as an abstract guide to what phase a learner might be at (Schmidt and Lee, 2014, p. 212). A large number of models have been proposed to describe the stages a learner goes through on the path from novice to master, but a useful and relevant example of these is Fitts’ stages of Learning which is presented below (Fitts and Posner, 1967).

Fitts’ Stages of Learning

Fitts proposes a 3 stage model of motor learning with an emphasis on both the perceptual and motor components.

Stage 1: The Cognitive Stage In this phase, gains in performance are rapid as the learner transitions from having no skill whatsoever to one where strategies are discovered to achieve the goal. Much of the learning is verbalisable, the focus is on goal identification, performance evaluation and questions around what to do, what not to do and how to achieve the aim. At this stage, instructions, demonstrations and video captures are very useful and any approaches that can leverage or draw attention to existing knowledge and chain it together to approximate the desired skill in this new situation will be valuable. A learner at this stage is also trying to grasp the environment: what to look at or look for; what to listen to or feel for when generating an appropriate movement is critical. It is acceptable if movements are not assured at this stage; it is a starting point for later performance gains.

Stage 2: Fixation Stage This stage is also referred to as the associative or motor stage and is much longer in duration than Stage 1. At this stage of learning, focus shifts to organising movements to produce the desired result more effectively. These refinements are based on the motor programs established earlier and for precise movements the learner develops ways to use the movement-produced feedback to guide their response.

Over the duration of this stage, performance increases steadily. However, there will be some inconsistencies as new approaches are auditioned but this decreases over time. Performers also become more smooth and assured in their movements and discover features of the environment that can be used to anticipate results and serve as cues. Crucially, they also begin to monitor their own feedback and detect their own errors.

Stage 3: Autonomous stage After a significant amount of practice the learner enters stage 3 and is capable of demonstrating expert performance. At this stage, the performer has a high perception of the environment which allows them to quickly and reliably extract information. Sequences of movement are developed and consolidated into courser grain motor programmes meaning that attentional demands are reduced and the performer can focus on strategy. Also, self confidence is increased and the performer has an increased capacity to detect and correct their own errors. However, improvements in performance are slower and marginal at this stage because the performer is already highly capable.

Fitts' model appears to be an appropriate conceptualisation of the learning process for dental operative skills due to its consideration of the wider cognitive and perceptual aspects that provide the executive control for determining the desired movements which are crucial for the safe provision of operative treatments.

3.2.3 Transfer of psychomotor skills

The learning process is of little value if as a result the learner can only perform the skill in the practice environment. Therefore, activities that facilitate psychomotor skills development have the ultimate aim of allowing the learner to generalise knowledge gained from the practice exercises so that they can be *transferred* to real-life scenarios. The study of transfer is how these skills acquired in one situation apply in another, however, it is frequently observed that this does not always happen (Singley and Anderson, 1989, p. 2). Therefore, it is of interest to educators to be able optimise learning activities to maximise the degree to which they impact upon later performance in the real world. Applied to dental education, this is the need to understand how skills developed in the simulation environment transfer to (or fail to transfer to) the clinical environment. There are a number of theories surrounding the transfer of psychomotor skills that will be discussed below.

Transfer from similar tasks

One of the oldest theories concerning the transfer of skills is the theory of 'identical elements' presented by Thorndike and Woodworth (1901) which theorises that the degree to which practiced skills will transfer to the target environment is based on the degree to which they share identical elements. However, what constitutes *identical* or is merely similar was not explicitly defined and led to the theory being criticised and questioned as to the extent the skills need to be identical. For example, 'if one learned to drive in nails with a yellow hammer, would these not transfer when using a red hammer?' (Singley and Anderson, 1989, p. 6) However, as discussed in Section 2.5 the idea that a simulation should be as realistic and as 'identical' to the real situation as possible has persisted and is perpetuated to this day.

More modern interpretations of similarity between tasks are more nuanced and similarities can be grouped into several classes of feature (Schmidt and Lee, 2014, p. 218):

- *Common movement patterning*: If there are common movements involved between two tasks then there is likely to be some degree of overlap. However, this only applies to movements that are within the same class of movements, for example throwing a ball two different distances. The theory suggests that if additional practice is given to one variant within the class then this should be transferable to others within the class.
- *Shared perceptual elements*: If similar perceptual skills are used between two tasks such as learning to track and intercept the flight of a tennis ball versus that of a football. Training exercises that improve one of these tasks are likely to be transferable to the other.

- *Shared strategic or conceptual elements*: Where there are strategic and conceptual similarities this permits transfer. Driving a car at different ends of the same country share strategic elements because the rules of the road are the same even though the roads themselves might be unknown.

Programmed Instruction

Skinner (1961) proposed that complex topics could be decomposed into a set of discrete successive steps or 'frames' and when material was provided in this way transfer would be improved through the provision of focussed reinforcement. Whilst focussing on transfer in general rather than motor skills specifically, this approach stresses the role of cumulative learning and feedback where the learner is permitted to master each of the frames before moving on to the next. These subtasks form a hierarchy and consequently form a series of 'learning sets' (Harlow, 1949): super/subordinate capabilities where successful completion of a task depends on mastering the skills below it. When transfer occurs between tasks which share similarities at a broadly equal level of difficulty, transfer between these is referred to as 'lateral transfer'. Whereas, when a more complex task depends upon mastery of a simpler task for progression, this is referred to as 'vertical transfer' (Singley and Anderson, 1989, p. 15). Whilst students may be able to grind out a solution to a complex task without mastering the preceding steps, Resnick (1975) showed that this is associated with greater variance between participants. However, this approach requires that the elements of the skill have been correctly identified and sequenced in a meaningful way, but when done so effectively learning rates can be faster by a factor of 2 (Singley and Anderson, 1989, p. 15). The decomposition of tasks and a mechanism to do so is explored in Chapter 5.

Analogical transfer

With analogical transfer, participants are reminded of a similar problem they already know how to solve and then are guided how the solution maps to the current problem, leading to improved transfer. However, learners often struggle to notice the similarities when drawing an analogous solution because the knowledge has been encoded in a different or distractor context (Gick and Holyoak, 1980) and can even sometimes struggle to remember how to perform the *same* task when it is presented in a different environment (Anderson et al., 1996).

If a candidate pilot were given a test that required the manipulation of a rudder and stick whilst monitoring an artificial horizon it would be seen to have clear face validity in determining the suitability of airline pilots. However, Fleishman (1975) showed that a test of keeping a stylus in contact with a rotating target on a device resembling a record player was equally valid as a predictor for selection. So as a tool for prediction, the face validity of the device used is not of immediate concern, but as a tool for the transfer of a cognitive skill the relevance must be apparent to the learner and pointing out similarities has a dramatic effect in their ability to apply the learning from another context and transfer it to the present task (Singley and Anderson, 1989, p. 221). Therefore, in order to be effective, abstract or analogous tasks should be explicitly linked to the problem domain so that learners are not forced to search for the skill's relevance (Gick and Holyoak, 1980) and tutors should point out the relevance of any pre-existing skills that can be applied in the new context (Anderson et al., 1996) so as to not risk overloading the learner's working memory with irrelevant details from the analogy context (Gick and Holyoak, 1980).

When done effectively the use of abstract training tasks have been seen to be highly efficient. Biederman and Shiffrar (1987) showed that abstract drawings of the discriminating features to be used when sexing day-old chicks from pictures could allow novices to quickly achieve accuracy results comparable to experts. Whilst the study did concede that the experts had reservations about the use of pictures due to the fact that some of the physical cues that were commonly used were not available, the fact that the novices were performing comparably to experts of many years experience so rapidly could suggest that abstract instruction can be a highly effective approach when learning these discriminating features. This study perhaps successfully performs the separation between the learning activity (the abstract drawings) with the test activity (sexing the chicks) discussed above and avoided the issues whereby goal attainment (successfully determining the sex) can interfere with the skill acquisition (knowledge of the discriminating features) such that the activity led to the desired learning for use in the transfer task rather than simply achieving the problem-goal (Sweller, 1988). This approach appears to maximise the generalisability of the skill during acquisition by optimising for 'transfer appropriate processing' allowing more efficient performance later (Salmoni, 1989). Furthermore, when abstract instruction such as this is presented alongside concrete illustrations then the effectiveness can be even greater than either approach alone (Anderson et al., 1996).

Specificity of transfer

Whilst the similarities between tasks do appear to have some impact, motor learning is quite specific. A learner tends to 'learn what they practice' and the sensory feedback (e.g. visual, tactile, auditory) received by the learner during specific types or locations of practice become part of the learner's representation of that skill and future performances are not as skilful when any of these feedback channels are altered (Proteau, 1992). This phenomena is recognised as the 'specificity principle' which begins to influence transfer as learning progresses (Henry, 1968). This principle suggests that as skills become more refined through practice, the movements associated between tasks which appear to be similar become more distinct and the differences between them increase. So, practice at a particular task (not of interest) because you would like it to transfer to another task (which is of interest) is not an effective use of time when it is considered that the time could have simply been spent practicing the task which is of interest (Schmidt and Lee, 2014, p. 218). Furthermore, there is a large and consistent body of literature that supports the conclusions that correlations between different skills are low and even skills that appear to be similar often correlate poorly and this lack of correlation argues against the concept of a general motor ability that can be increased in the abstract (Schmidt and Lee, 2014, p. 158).

This principle also applies to 'lead-up' activities such as suturing a grape, where this is a cost-effective approach when compared to a more expensive simulator or the use of real tissue. However, the context has a pervasive influence on retention (Shea and Zimny, 1983) and these kind of activities tend to transfer in line with the similarity with the target task itself (Schmidt and Lee, 2014, p. 218). Therefore, practice in simulated environments must recognise the specificity of learning as a dominant characteristic in order to be effective (Schmidt and Lee, 2014, p. 200). The presence of transfer between similar tasks leads to a common misconception that abilities can be improved and trained through drills or other activities that improve some shared underlying ability. This has an intuitive logic, but generally they do not work because the generalised skill has become specialised by being interpreted in a domain-specific way (Abernethy and Wood, 2001) so an attempt to increase underlying abilities via a non-specific practice task is usually ineffective and only serves to increase competence at the practice drill rather than transfer to the target task itself. Again, the time would be better spent practicing the target task itself.

However, some tasks are complex and attempting to learn and practice them all at once would overwhelm the student. As discussed above, a common approach to overcome this is to divide the task into smaller chunks for part-task practice and then as the sub-tasks are learned they are re-integrated into the whole skill later. Where this approach is chosen, questions arise as to how to decompose the sub-tasks such that they can be practiced to maximise the transfer to the target skill. Some tasks, especially ballistic actions, can't easily be decomposed and practicing individual components can lead to a worse performance than not practicing at all (Schmidt and Lee, 2014, p. 220). This is because practicing the sub-task has different characteristics when carried out in isolation and effectively becomes a distinct and separate skill from when it is deployed as a whole. However, this doesn't preclude its usefulness early in training, because if there are a lot of steps to perform part-practice might help by presenting the parts separately then grouping them in to progressively bigger chunks to build up to the whole task as quickly as possible.

Transfer of skills acquired using VR

The above discussion leads to important considerations in the use of VR for dental education and the degree to which skills learned in a VR environment will transfer to the operative context. As discussed in Section 2.4 the hardware of the simulators is unable to faithfully and accurately recreate the subtle tactile feedback found when performing the procedure in the real world. Consequently, the tactile sensations encountered in the VR environment are different to that which would be encountered in the real world. The specificity principle could imply that the use of these simulators to develop fine motor skills could be ineffective and would transfer poorly to the real context because the differences mean that the learner is (in effect) developing a different skill: one of being proficient on the simulator.

Fortunately for proponents of the use of VR in dental education, transfer is not restricted to fine finger dexterity. The declarative representation of the rules that the learner is applying to the task are more abstract, so any corrections and insight gained in the the simulation context can be equally applied to the representation in a different context (Singley and Anderson, 1989, p. 193). Therefore, these can be developed and refined even if the fine finger dexterity itself is not transferred. Additionally, there is evidence that students learn new knowledge more effectively in courses where there is active, inquiry-based and collaborative learning using information technology (Hake, 2007). When this is coupled with observations that students have a positive opinion of the use of VR in dental education (Ben Gal et al., 2011) and engagement also correlates with an increase in subject-specific learning (Kuh, 2003), perhaps using VR with a focus on the declarative aspects of operative skill would be an effective way of balancing these considerations. This approach is explored in detail in Chapter 8.

3.2.4 Schema Theory

It is now worth considering what mechanisms a performer uses when generalising skills from practice. Schema Theory can trace its roots back over 100 years across many authors and it asserts that knowledge is stored in long term memory using 'schemas' (Sweller et al., 1998). An individual schema can be considered as a set of rules for determining if a stimulus is part of a pre-established category and is encoded in the manner in which it will be used (Schmidt, 1975). For example, if presented with an image of a dog, this doesn't need to be a dog (or even a breed of dog) that has been seen before, it is simply evaluated in terms of previous experience of its 'dogishness' to establish if it should be categorised as a dog. Schemas form the basis of knowledge; and expertise is built through the acquisition of increasing numbers of schemas and through the combination of lower-level schemas into progressively more complex representations (Sweller et al., 1998).

Schema Theory was extended to cover motor learning by Schmidt (1975). Prior work by Adams (1971) had suggested that a specific motor program was required for each movement pattern and that these would gain strength and become more refined if practised. However, this could present a storage problem if every new movement pattern was required to be stored in memory independently, nor does it explain how skills can be adapted to new situations. Therefore, Schmidt (1975) suggested the presence of a more abstract (or general) motor program template for a skill which accepts parameters (such as speed, distance, current body position etc.) to tailor it to the novel situation presented. Expertise is linked to these schemas such that an expert performer has encountered more problem scenarios and therefore has more parameterisations of the skill stored, meaning, that experts are able to recognise (or adapt to) the configuration of the present situation and deploy an appropriate pre-existing schema to solve the problem. However, a novice without this library of schemas must deploy less efficient general purpose problem solving strategies (with the associated additional effort required) such as a means-ends approach to find a candidate solution or novel parameterisation of an existing schema (Sweller, 1988).

Physical actions can be classified as being under either open-loop or closed-loop control. Open-loop controlled actions are rapid movements such as throwing a dart or kicking a ball and are executed subconsciously without reference to feedback during the motion. These actions can be thought of as rehearsed 'phrases' of a vocabulary of actions that have developed through practice so that when a stimulus is encountered then the motor instructions are sent to the muscles without conscious oversight. Conversely, closed-loop control is a system where adjustments are continually made by feeding output information back in as an input. Here, when the stimulus for an action is encountered the action is carried out, but kinesthesis and proprioception allow the performer to tell if the execution of the task 'feels' right relative to an internal reference of correctness and then feed that information back in to the performance of the task allowing for correction of errors (Schmidt, 1975). This fine-tuning (or specific parameterisation of the schema) can then be stored for future use increasing the skill level. In order for the skill to develop the performer needs to store four things:

- The initial conditions: such as their position in space, the environment, visual/auditory cues etc.
- The specifications or adaptations chosen for the general motor program to be executed to achieve the goal.
- The sensory feedback from making the movement, such as: how it felt, what it sounded like, what was seen etc.
- The outcome of the movement in terms of whether it achieved the goal or what the deviation from the desired result was.

These four sources of information are then stored together (Schmidt, 1975) and as more attempts are undertaken the relationships between them are abstracted into a generalised motor program that can be applied with appropriate parameters to novel instances of that class of motions. Many dental operative skills would be regarded as closed-loop activities because, for example, the sensations detected via instruments are used to inform decision making and diagnosis.

Key to the development of these motor schema is knowledge of the result (often abbreviated to KR) because the input specifications are evaluated in terms of their correctness based on the outcome (Schmidt, 1975) so it is important that the performer knows what the desired outcome is and if they achieved it. If the parameters chosen produce the desired result then these choices are strengthened and will be more likely chosen next time. More so, error detection increases with practice and if sufficient feedback is provided the subject is able to continue to learn once the feedback is removed because the subjective reinforcement allows recognition of errors (Schmidt, 1975).

However, further to the discussion above, this suggests that current VR simulators may not be appropriate to develop the schemas associated with operative performance because the haptic feedback does not match what will be encountered in the real world. With their present limitations, when a successful performance is produced in VR, this strengthens the schema that produced that motion. However, in the real world different stimuli are received and because of the specificity of transfer, the parameterisation will not be triggered because it is effectively (at the level of encoding) a different skill. Similarly, this may go some way to explain why there was little difference detected in the performance between novices and experts at VR exercises in Wang et al. (2015a) or that the performances rapidly converged to expert standards in Suebnukarn et al. (2014). The differences in sensations that the VR simulator was providing may mean that the experts and the novices were *both* encountering a new and novel situation. It may be that the experts were calling on a previously un-encountered extreme parameterisation, if not different altogether, series of motor programs in the performance of the task and when the expertise reversal effect is factored in, compounding the differences experienced by the expert performer, it suggests why their performance might be reduced to that of a novice. This is not to suggest that simulation and simulated practice are not valuable for a learner, but that improvements seen as a result of simulated practice may be due to a better understanding of the task rather than refinement of the performer's finger dexterity directly.

3.2.5 Cognitive Load and the structure of practice

A learner can only monitor and work with information that they are conscious of so in order to develop a schema it must first be processed in working memory (Sweller et al., 1998). Cognitive Load theory builds upon schema theory and is based on the central assumption that there are a finite number of items that working memory can attend to of seven plus or minus two (Miller, 1994). As domain specific knowledge is learned over time it becomes encoded in a schema and stored in long term memory. Importantly, the 7 ± 2 limitation only applies to novel information, so schemas retrieved from long-term memory do not contribute towards this limit. Consequently, the theory proposes (and has been shown to be true in a variety of domains (Sweller et al., 1998)) that the presence of a large number of pre-existing schemas (rather than simply being good at problem solving) is the primary distinguishing factor between experts and novices (Sweller, 1988). Therefore, the development of expertise is linked to the presence of pre-existing schemas and not from the ability to hold and manipulate large amounts of new and novel information (Merriënboer and Sweller, 2010).

However, in order for a new experience to transition from working memory to long-term memory (and develop expertise) the novice learner must be able to construct an appropriate schema for long term storage. The ease with which they are able to do this is impacted by the cognitive load of the task. The total cognitive load for a task is made up of three elements: intrinsic load, extraneous load and germane load (Merriënboer and Sweller, 2010).

The intrinsic load for a task is its inherent complexity and it can't be changed without changing the task itself (e.g. through simplifying or decomposing it into a smaller task). It is linked to the number of interacting elements that can't be understood individually so must be held and processed in memory all at once. Tasks with a greater number of elements that must be held in memory and processed at the same time are more difficult than those with fewer (or those where elements can be mentally separated) and accuracy decreases when judging multiple attributes simultaneously (Miller, 1994). As expertise develops, the intrinsic cognitive load of the task decreases as, what were previously individual interacting elements, become encoded as a single coarser-grain schema. This long-term memory schema no longer impacts upon the available cognitive resource and makes the task less taxing cognitively.

Extraneous load is caused by the method of instruction. If a learner is not provided with guidance, has to integrate information from various sources or is unclear of what defines a successful performance then the extraneous load is increased. More bluntly, it is the overhead of effort required to process poorly designed instruction (Sweller et al., 1998). This increases the difficulty in accomplishing the intended learning, but it is the factor that is partially under the control of the instructor to address through more efficient teaching approaches. For example, working memory is thought to be divided in to a 'visual-spatial scratchpad' and a 'phonological loop' to deal with visual and auditory information respectively (Sweller et al., 1998). Within limits, these channels can be utilised independently to increase working memory capacity so well designed visual information accompanied by relevant auditory narrative will not impose the same load as presenting the same information via a single channel.

Finally, germane load represents the cognitive resource required for organising and processing the information contained within the intrinsic load which leads to the construction of schemas; effectively germane load is the effort required for learning to take place (Sweller et al., 1998). It is important to note that intrinsic and extraneous cognitive loads are cumulative. If the sum of these loads is too great then the learner will become overloaded and there will be no cognitive resource remaining to be utilised for the learning to take place (Merriënboer and Sweller, 2010). Therefore, instructors should take steps to reduce the extraneous load to make the tasks more manageable, but also by considering the germane cognitive load, instructors can assist the learner in the construction of these schemas and facilitate their transition in to long term memory. Once established in long term memory, the schema can be re-used without impacting upon working memory meaning that more resource is available for future learning, comprehending a new and novel situation or deciding upon an appropriate response.

The structure of practice sessions

As discussed above, controlling the cognitive load of the task can have a positive impact upon the acquisition of a skill, but the way that the practice sessions themselves are structured can also impact upon how effectively and efficiently a learner is able to establish the parameters of the target skill. Expertise is based on the acquisition of increasing numbers of pre-encountered parameterisations of schemas and there is considerable evidence that varying the values of these parameters during practice facilitates more effective skills development (Schmidt and Lee, 2014, p. 243) than simple repeated practice at the target skill. Perhaps intuitively, learners who exclusively practice at a single member of a class of target skill will perform better than those who practice a variety of variations of the class when faced with the task they have practiced. However, they will perform more poorly when presented with a novel example. The variable practice allows the skill to become more generalisable so that learners are better equipped to deal with unseen instances of it (Schmidt and Lee, 2014, p. 243).

Additionally, mixing the order of tasks can also have a positive effect on learning. For example, if a learner is practicing three completely distinct tasks: performing them in blocked-practice where the learner practices all A, then all B followed by all C results in a better performance during practice, but a poorer acquisition than randomly practicing them. This has two surprising consequences: firstly it would be expected that focussing on an individual skill allows refinement and correction, and secondly, it is not unreasonable to associate better performance when practicing to better performance when deploying the skill. However, neither of these appear to be the case and two possible explanations have been identified. Shea and Zimny (1983) presented the 'elaboration hypothesis' such that the random order made the practice tasks more distinct so participants drew out analogies, relationships and other distinctive features. However, the blocked practice allowed the learner to switch off and not engage the higher level of thinking or perhaps an 'Einstellung Effect' occurred where the learner became mechanical in the application of the methods and began to assume that all problems are of the same type (Luchins, 1942). However, Lee and Magill (1983) suggested that it was caused by the fact that Task A was in short term memory which was lost during Task B. This meant that the representation had to be re-constructed each time the task changed. Doing so over a number of trials led to the construction becoming more efficient; this has become known as the 'forgetting hypothesis'.

VR simulation has the potential to present the learner with a vast library of different tasks with randomisation and variation in the parameterisations and the order in which they are presented. This would help to ensure that the student is faced with a diverse range of challenges in order to develop the broad base of experience to build the schemas necessary for expertise. Equally, VR is a highly controllable environment where prompts and informational feedback can be provided automatically at the most appropriate time. These factors could combine to make VR a highly effective environment in which to develop psychomotor skills but, as discussed above, the limitations of the hardware mean that care must be taken to ensure that the desired outcomes and improvements from VR practice align with what is likely and possible in this environment.

3.2.6 Overview

Much of the psychomotor literature discussed above is based upon the development and measurement of a skill in a more closed system. The research that supports these theories frequently originates from sports science, psychology or mathematics education where the tasks can be contrived to result in a closed system with defined goals for measurement. For example, developing skills in sports science can measure the accuracy of a throw of a ball into a net, the swing of a club to hit a ball into a hole etc. However, operative dentistry is much more open; there isn't, for example, a single cavity to get good at cutting. Beyond generalisable shared attributes, every instance the learner will encounter will be unique and unseen. By the time a student is safe to operate on real patients, their motor skills are arguably already at the maintenance phase (Chambers, 2012) and it would likely be unethical for them to operate if they were not. Therefore, it is arguable that instead of always focussing on improving fine motor skills, the focus of at least some practice sessions should be focussed on preparing the student to contend with this unpredictability and reliably apply appropriate principles to unseen scenarios.

Through the lens of the psychomotor skills literature, the differences between phantom head and VR practice at an operative procedure becomes more pronounced. Even though typodonts, as commonly used for practice, do not feel the same as cutting real tissue, the tools themselves are real and the setting has many similarities to the clinical environment. Schema theory suggests that the difference in density found in typodonts could be accommodated as a parameterisation of the skill. However, the equivalent practice in a VR environment may not lead to the same development of transferable skill because not only is the tooth not real, the behaviour of the cutting instrument is not real, nor does the more ‘IT suite’ situation necessary for electrical devices provide similar environmental cues. Perhaps, the differences are simply too great to reliably develop transferable finger dexterity.

However, there is a link between a learner’s actions and the underlying declarative knowledge that prompted that response. If the underlying knowledge is shallow or ‘rote’ then the responses will likely have the same character (Singley and Anderson, 1989, p. 2). If this declarative knowledge or “insight in to the nature of the skill” (McCloy, 1968) is the foundation upon which the ability to transfer uses of that knowledge is based (Schmidt and Lee, 2014, p. 218), and a self awareness of performance correlates most strongly with pre-clinical performance (Lee et al., 2017) then perhaps focussing on developing the cognitive aspect of the skill rather than finger dexterity in a VR environment would lead to the greatest gain? After all, much of medical education is founded on the assumption that any increased knowledge and skill acquisition will eventually lead to an improvement in patient care (Colt et al., 2011) so investigating how VR can be used to increase the attributes that most closely correlate with improved performance would likely be the most valuable use of the technology.

Chapter 4

The Research Question

The design requirements for early VR simulators were framed by traditional expectations of phantom head based instruction. There was an expectation that VR dental simulators could be developed such that they could be used in the same way and teach the same skills as traditional phantom heads. However, the different modes of interaction shifted staff and student attitudes towards these new devices and their perceived value (San Diego et al., 2012). This early desire to create a realistic simulation led to it being the most common use case for VR simulation in dental education reported in the literature.

It is often suggested that VR simulation is preferable to traditional phantom heads as it overcomes many of their issues (Sofronia et al., 2012; Eve et al., 2014; Ruthenbeck and Reynolds, 2015) but they are often criticised for not being 'realistic' enough (Tse et al., 2010; Konukseven et al., 2010; Ben Gal et al., 2011; Eve et al., 2014). The perception of realism is a complex and multi-facetted concept, cutting across the software, hardware and how the user perceives and interacts with them. From the software perspective there are questions around the the visual realism – how realistic does the tooth being operated on or the overall scene look? the behavioural realism – does the subject react to interactions in the way expected e.g. do removed particles just disappear or are realistic debris shown? the aural realism – does it sound realistic (and does the sound react to changes in the environment such as when cutting through different materials)? Similarly, for the hardware, does the haptic device provide sufficient tactile feedback to be considered realistic? Finally, when the user interacts with the system, does the configuration produce sufficient environmental realism to allow the learner to adopt a realistic posture and interact with the tools in a way that is similar to the real environment?

Within these issues, there are unexplored subtleties that impact upon the authenticity of simulation: does the device held by the user need to represent the real tool being used in the simulation? Does the visual realism need to be photo-realistic or is a reduced almost 'cartoonish' appearance adequate? Could a lower fidelity visual representation have higher authenticity (Gilbert, 2016) and thus greater educational value? When scoring a user's performance is the scoring mechanism robust enough for the intended purpose (Messick, 1989), is it validated for it (Wang et al., 2016b; Messick, 1995) and can a user improve their technique using the scores provided (Larson et al., 2013)?

Many of these issues can trace their origin back to the desire to create a digital phantom head and reduce the need for the tutor's presence. By simply trying to replicate traditional ways of teaching using VR, it shows itself to be a representational, incomplete facsimile of a phantom head (San Diego et al., 2012). In this context can VR *ever* be real enough? The simulation provided by a phantom head is much closer to the real situation by virtue of using real tools that provide the exact same sensations as will be encountered operatively. As a student practices, auditioning different motions and varying their positioning, the chosen parameters for the underlying schema that led to the most successful performance will be strengthened (Schmidt, 1975) and can be re-applied in the target context. However, with the differences in sensation that the VR simulators provide, the specificity principle (Henry, 1968) suggests that the learner will be less skilful when applying them in the target context (Proteau, 1992) and practice on these devices merely serves to increase their ability at VR simulation rather than their actual operative skills.

However, the very attributes that suggest traditional simulation is a better tool to enable transfer to the real-world context are the same ones that reduce its effectiveness as a learning tool. Contending with a realistic real-world environment and the accompanying additional stimuli increases the intrinsic cognitive load of the task adding to the difficulty for the learner to identify and associate the correct response with a given challenge (Hammerton, 1967). Working memory struggles to hold and reason with this volume of complex interacting information whilst applying it to a training task, and any training environment that asks learners to do so is not likely to be optimal (Sweller et al., 1998). Considering these factors, attempting to use VR to fulfil the exact same educational role that traditional simulation already occupies does not appear to be their best use in pre-clinical education.

For success on a dental degree, students are expected to integrate skills across the affective, cognitive and psychomotor domain (Segura et al., 2018). Phantom heads are effective at providing an environment to detect and refine the subtle tactile sensations to develop psychomotor skills, but incur additional cognitive load which may impede learning. VR dental simulators can provide an environment where the cognitive load is managed, but are likely to be less effective for developing transferable finger dexterity skills. There is no reason that the two simulation modalities cannot co-exist and each serve different goals. VR simulators could offer a focus on the cognitive aspect, providing a unique opportunity to develop skills in a way that would be difficult or impractical to achieve in a traditional environment and the phantom head environment can focus on the development of the psychomotor aspect allowing the learner to demonstrate the cognitive knowledge acquired from VR.

Traditional tuition in a phantom head environment encourages a learner to attempt operative procedures using skills acquired from tutor demonstrations, printed material and instructional videos (Field et al., 2020). However, words are a crude mechanism for describing movements and any complex movement (for example tying shoe-laces) is very difficult to verbalise (Schmidt and Lee, 2014, p. 231). Books use the written word so suffer some of the same verbalisation limitations. Pictures are static and to provide multi-stepped instructions often require instructions to be overlaid on the image, these need to be integrated and because they both rely on the visual-spatial scratchpad place additional load on the learner (Sweller et al., 1998). This means that much of the responsibility for the instruction falls on instructional videos and tutor's demonstrations. Whilst videos are an improvement on text and static pictures, they not interactive and passive observation is insufficient (Diamond, 1996). Demonstrations are an effective tool, but place a lot of the outcome on the skill of the demonstrator and are somewhat limited by the effective size of group that can observe. Theoretically, VR could overcome these and offer a realistic environment where these skills can be taught and demonstrated, but the limitations discussed previously suggest that these systems are not at their most effective when attempting to recreate a realistic environment. However, exercises which are based on analogical or abstract transfer have been seen to be effective (Biederman and Shiffrar, 1987; Gick and Holyoak, 1980) and providing exercises in this form would avoid the limitations and complications of a realistic simulation. Additionally, by not trying to recreate the holistic environment as seen in the real world, the cognitive aspects of the task could be decomposed in to simplified 'frames' (Skinner, 1961) to aid learning. These frames could be implemented as focussed 'deconstructed' exercises. Concentrating on individual concepts could reduce element interactivity and allow them to be learned serially rather than simultaneously (Sweller et al., 1998) and the knowledge more readily re-constructed by the learner. These exercises would be interactive and the information would be directly relevant to the task (Pickering, 2017) potentially showing a superior signal-to-noise ratio (Gilbert, 2016). These features could facilitate the germane cognitive load necessary for the learner to efficiently construct schemas based on the declarative knowledge rather than simply hoping that learning will occur as a result of high-fidelity simulation (Salas et al., 1998). This knowledge could then be recalled from long term memory when applying it in a phantom head environment, freeing up cognitive resource to permit levels of performance and enhanced learning in the subsequent context that would not otherwise be possible (Sweller et al., 1998).

This abstract part-task approach to the use of VR should allow the learner to develop a broad declarative base of knowledge that is more readily transferable to the target environment, result in superior retention, a deeper understanding and serving as a better vehicle to transfer the knowledge in to practice (Zigmont et al., 2011) than the current approach of whole-task simulation.

The remainder of this thesis will explore:

> **What is the value of deconstructed part-task training for learning and teaching pre-clinical operative dentistry with VR simulation?**

4.0.1 Hypotheses

- H_1 – A participant who receives instruction based on the cognitive aspects of an operative task using a series of 'deconstructed' part-task exercises on a VR simulator will be able to apply procedural aspects of that knowledge in a transfer task more accurately than one who was taught using a whole-task VR simulation.
- H_2 – A participant who receives instruction based on the cognitive aspects of an operative task using a series of 'deconstructed' part-task exercises on a VR simulator will have superior retention of the declarative knowledge in a transfer test than one who was taught using a whole-task VR simulation.
- H_3 – A deeper understanding of the cognitive aspects of the task acquired from instruction using 'deconstructed' part-task exercises on a VR simulator will allow participants to make better judgements, closer to those of a qualified practitioner, in a transfer task than those instructed using a whole-task VR simulation.
- H_4 – VR exercises based on deconstructed tasks will have a lower face-validity than whole-task and be less well received by novice learners
- H_0 – The Null Hypothesis for $H_{1\text{-}4}$ respectively is that the intervention received (deconstructed or whole-task) is independent of performance at the transfer tests.

4.0.2 Aim

The aim of this project is to assess if Virtual Reality, as an operative dental education tool, results in superior learning (as measured via a transfer task) when it is employed to provide deconstructed part-task exercises instead of facilitating whole-task repetitive practice.

4.0.3 Objectives

- Establish a process through which a task can be deconstructed into its sub-component tasks and reveal its underlying cognitive base.
- Identify an appropriate dental operative procedure that is currently taught using a whole-task approach with VR simulation.

- Apply the process to deconstruct the operative procedure and identify appropriate sub-components for educational exercises
- Develop relevant exercises for the identified sub-components of the procedure for execution on the VR simulators available at the host institution
- Run a pedagogic randomised cohort experimental trial using junior students where:
  - Experimental Group: Receives instruction using the newly developed 'deconstructed' part-task exercises
  - Treatment As Usual Group: Receives instruction using a whole-task VR simulation exercise
  - Performance of the two groups is compared using a number of quantitative measures of performance, retention and transfer to establish superiority
  - Gather participant opinion on the intervention exercises to explore any themes relating to their perceptions of the approach and its acceptability

This approach frames the use of VR within Bloom and Krathwohl's (1956) cognitive domain. The part-practice exercises focus the intended learning on the stimulus identification and response selection phases of the information processing model (Schmidt and Lee, 2014, p. 22), aligning with Fitts' Stage 1. Whilst outside the scope of this work, this use would be intended to form part of a broader educational approach whereby traditional simulation leverages the cognitive skills developed in VR to develop Bloom and Krathwohl's (1956) psychomotor domain skills and Schmidt and Lee's (2014) movement programming stage. The phantom head training would then be employed to address Fitts' Stage 2 into Stage 3 as the student transitions to the clinical environment. Finally, in the clinical environment the student would demonstrate the skills acquired from the two simulation modalities, focussing on final refinements of their psychomotor skills and the interpersonal attributes required for Bloom and Krathwohl's (1956) affective domain.

This work is intended inform a useful role for VR simulation in the curriculum by not trying to replace phantom head simulation, and instead providing students with a strong transferable declarative base of knowledge and the tools with which to interrogate and reflect upon their own performance when deploying the skill. This approach may lead to improved performance and ultimately an improved patient experience.

Chapter 5

Task Analysis of a Dental Operative Task

This chapter describes the development of a novel approach to the deconstruction of operative tasks based on Task Analysis methods which is then applied to describe the procedure of caries removal and cavity preparation. Finally, the method used is discussed and evaluated before the resultant Task Description is further explored in Chapter 6.

5.1 Defining task analysis

Task Analysis is a process which can be applied to tasks in order to examine them in terms of the actions and/or cognitive processes performed by operators in the pursuit of some goal (Kirwan and Ainsworth, 1992). It is an iterative process that involves preparation, data gathering, organisation, analysis and recommending solutions based on the findings (Adams et al., 2013). When used for instructional design the nature of the task that the learners will be performing must be understood (Jonassen et al., 1999, p. 3). Therefore, task analysis should precede other stages in the design process as it is responsible for the identification of what the learner is expected to learn. This is achieved through careful analysis of the task's underlying processes and concepts so that detailed, unambiguous and effective ways of teaching them can be specified (Miller, 1953; Annett et al., 1971; Resnick, 1975; Jonassen et al., 1999; Mannan, 2005).

Task Analysis is domain independent (Onnasch et al., 2016) and has been applied across a diverse range of industries and contexts: from its original association with a military context (Drury, 1983), to safety-critical human factors in the process industry (Mannan, 2005), business process analysis (Papavassiliou and Mentzas, 2003), organisational knowledge capture (Basque et al., 2014), aviation (Onnasch et al., 2016), medical

(Rosen et al., 2002; Tjiam et al., 2012), dental (Walker and von Bergmann, 2015) and even to identify the task analysis process of the process of analysing tasks (Adams et al., 2013). Underlying these diverse uses is the common theme that efficiency, safety and productivity gains can be derived from a recognition and improvement of instruction where users interact with systems (Kirwan and Ainsworth, 1992).

Instruction is rarely complete, and as a result, the simplifying and organisational principles that underly the material being taught are often left to the learner to discover for themselves (Resnick, 1975). Tasks are perceived in segments and larger tasks are implicitly broken down or *decomposed* by learners (Zacks et al., 2001) so failure to address this at an instructional design level could introduce extraneous cognitive load and lead to less optimal outcomes. Whilst some learners are able to independently derive the underlying structures, others require explicit help to find efficient strategies for performance (Resnick, 1975) and the strategies they used to break the task down may be of variable effectiveness (Cheng et al., 2015). If the instructor assumes the responsibility for decomposing the task, the strategy is enforced by the approach (Cheng et al., 2015) and the method of instruction scaffolds the knowledge on behalf of the learner. Therefore, for effective instruction, it is essential to understand the nature of the task that learners will be performing in order to guide them to the desired ways of thinking and acting (Jonassen et al., 1999). Task analysis can produce such an understanding through the production of a *blueprint* for instruction which mitigates the risk that important content and activities present in expert performance are missed or not understandably linked in teaching materials (Jonassen et al., 1999, p. vii).

Modelling of pre-existing knowledge to produce a task analysis, whilst cognitively demanding, is of value in itself. It stimulates discussions, aids the conceptualisation of a task and, allows experts to identify and uncover knowledge that is central to their expertise that they may have previously considered incidental or trivial (Basque et al., 2014). This recognition allows for improvements in teaching, because instructors are now aware that they must now convey the missed information and can deploy the most effective instructional approach to do so (Schmidt and Lee, 2014, p.162). Instruction based on task analysis, where the resultant tasks are trained in their component form, have been shown to have advantages over practice of a single complex task. They show higher levels of performance on the criterion task for the same amount of training time, compensate for lower (initial) ability and develop more readily transferable knowledge and skills to other contexts (Frederiksen and White, 1989). During the training sessions themselves, whilst initially taking longer to complete, the smaller tasks result in fewer mistakes, are easier to perform and performance is more resilient and consistent in the face of interruptions (Charness and Campbell, 1988; Cheng et al., 2015). Furthermore, the decomposed elements facilitate effective chaining together of steps into a smooth procedure (Charness and Campbell, 1988).

5.1.1 Task analysis in instructional design

Task analysis has been part of training and system design for over 100 years (Militello and Hutton, 1998) and, because it is needed whenever performance is analysed in component parts, many investigations in psychology have been based on task analysis even if the researchers weren't consciously aware of it (Resnick, 1975).

One of the key focuses for instruction is to develop problem environments from which the learner can acquire the desired knowledge (Frederiksen and White, 1989) because in realistic whole-task environments the signal-to-noise ratio is reduced such that the intended learning can be lost amongst many distractions (Gilbert, 2016). Furthermore, a theory of instruction requires a description of the states of intellectual competence and the relationships between them (Resnick, 1975). An important step towards this is the creation of sequences of problems and exercises that promote the development of these competences within the learner (Frederiksen and White, 1989). So, by developing exercises that focus students' attention on certain aspects of the problem structure, the likelihood of them acquiring insight is increased (Wertheimer, 1961). However, concepts are often interlinked: some need to be combined to produce an intended outcome and others depend on mastery of a dependent skill where not doing so will lead to failure at the target task (Resnick, 1975).

Task analysis provides structure to a learning intervention. It assists instructors and learners to recognise the nature of the problem and break it down in to more manageable sub-tasks. Through the recognition of the individual components it also allows specification of what knowledge and the level of ability a learner should be able to demonstrate in a given context (Paquette et al., 2006) leading to more elegant methods of learning and assessment (Resnick, 1975). However, this also leads to a question of how small individual components should be? In the extreme, procedures could be reduced to individual muscle twitches (Annett et al., 1971), however, more commonly the decision is based upon the formula: $P \times C$; when the cost (C) of failing at a sub-task multiplied by the probability (P) of a mistake occurring approaches zero (Mannan, 2005; Annett et al., 1971). However, the value of P is subjective and not every learner is of equal ability so the probability must include a lowest likely common denominator (Annett et al., 1971).

5.1.2 Limitations of task analysis

Unfortunately, task analysis is not a complete solution to the problem of developing instructional content. It can only provide an iterative process to assist highlighting problems and requirements that can point a skilled analyst in the direction of potential solutions (Drury, 1983). Whilst good instructional design depends on task analysis, task analysis does not ensure good instruction (even if the task analysis was performed competently) (Jonassen et al., 1999, p. vii).

A crucial limitation of task analysis is that highly proficient performers often do not know how they do what they do. With expertise, the execution of the movements becomes non-conscious (Schmidt and Lee, 2014, p.162) and encapsulated into a procedural ability so the skilled performer is not always able to explain what they are doing or identify the points where they are making decisions (Tjiam et al., 2012). The tacit knowledge becomes difficult to externalise because it is used 'live' without conscious recognition of what is being done and why (Basque et al., 2014). Whilst task analysis aims to unpack complex tasks so that instruction can be derived from it, it cannot be overlooked that many studies have shown the difficulties experts have in formulating detailed explanations of a task even when they are aware the explanation will be used by a novice (Basque et al., 2014). Nor can it be overlooked task analysis itself is an art and is dependent on the skill of the task analyst (Jonassen et al., 1999, p. 4) and the approach selected (Jonassen et al., 1999, p. 6) so different recommendations may be derived from the same source information.

The goals determined at the outset of the task analysis can also influence the results and recommendations. Rosen et al. (2002) classified motions based on video recordings and force/torque readings of a surgical procedure, describing these as the *prime elements* of tool/tissue interactions that would be inherent regardless of the modality. This is a level of analysis not far from decomposing a task all the way down to the individual muscle twitches as suggested by Annett et al. (1971) and a classification in these terms would likely result in educational material that instructs a learner to mimic the motions described. However, the specificity of motor skills discussed in the previous chapter may suggest that these *prime elements* are not as inherent and immutable as might be hoped and consequently may not lead to the optimal focus for the instruction. Others have taken analysis in different directions; to simply gain expertise and identify wider transferable skills (Frederiksen and White, 1989); to identify tasks with the highest cognitive load (Walker and von Bergmann, 2015); to identify 'decision points' within a task (Tjiam et al., 2012); or to identify the demands that are placed on an operator (Drury, 1983). These goals and areas of focus will influence the analysis such that, for example, if the goal is to identify task elements with the highest cognitive load, then the likely outcome from this analysis are recommendations on approaches to managing that cognitive load.

Another limitation when using task analysis to develop instructional material is that the patterns of abilities change with practice. A beginner must expend considerable cognitive resource on deciding what to do, remembering what order to do things in and to figure out the criteria for success (Schmidt and Lee, 2014, p.165). Novices are, by definition, still learning and this will be associated with weaker manual skills, a lack of familiarity with the procedure and each step may require conscious and deliberate decision making (Tjiam et al., 2012). Therefore, any description of the task must recognise the stage of competence of the intended audience and describe the characteristics in such a way that instruction can be arranged as activities to assist with the acquisition of skill, recognising that this may differ from the arrangement which is most efficient for expert use of that knowledge (Resnick, 1975).

It also must be recognised that experts make different mistakes, have a different cognitive load and as a result may not recognise the areas where novices might struggle. This is known as an "expert blind spot" (Walker and von Bergmann, 2015). For example, the angulation of the mirror is second nature to an experienced dental tutor, however, it is a challenge for novices (Walker and von Bergmann, 2015). If a tutor does not recognise this aspect is a challenge, they may not direct their feedback towards addressing it. Even the areas and rules that the expert describes can vary in complexity and may not be consistent, rational or optimal (Annett et al., 1971). Tutors also have a tendency to assume that learners are "like they were", but not everyone learns in the same way, at the same speed or with the same motivation (Walker and von Bergmann, 2015). These difficulties of identifying where novices struggle has led to some to involve student participants in the task analysis itself, however, this can introduce different issues stemming from their lack of experience that may unduly influence the analysis. For example, searching for the descriptive words and terminology puts a higher cognitive load on a student when describing a task (for example the name of a tool or the name for the surface of a tooth) so this can mean that a student is not struggling with the task itself, simply they are struggling to remember the term to describe it (Tjiam et al., 2012).

The above suggests that asking experts to participate in a task analysis is not without its limitations. However, the task analysis itself is only part of an overarching process. As discussed, the task description does not take into account the ability of the learner so directly teaching the routines uncovered by task analysis would be a mistake (Resnick, 1975). Learners reframe and adapt instructional content in order to find more efficient routines (Resnick, 1975) and a learner developing a motor skill must, at the very least, reframe the declarative instruction to perform it using their own limbs and specific abilities of finger dexterity (as compared with those of their tutor's) so truly direct teaching is by this definition impossible.

To develop generic transferable knowledge, the skills must be learned separately from their application through a series of training tasks. Some of these training tasks may be developed purely to train others (which may not ever be used in reality) (Frederiksen and White, 1989) but care must be taken to ensure that the training tasks do not hinder or confuse the acquisition of another. However, sometimes the whole task is more than the sum of its parts and these can't be developed out of context. A part-task training approach risks difficulties when transferring the skill or could be missing crucial components that neglect the strategic character of the integration of the subtasks (Frederiksen and White, 1989). Simply playing games that implicitly embody principles does not necessarily lead to learning of these concepts (Frederiksen and White, 1989), so even after a careful and competent analysis of the task, the development of training exercises is very much a matter of "artful development" (Resnick, 1975). It is the analyst's responsibility to record what seem to be the most promising training possibilities to enable a novice to learn how to select the most appropriate solution to a given task and then to represent it in a form from which a student can learn (Annett et al., 1971).

5.2 Identification of a dental operative task for analysis

As discussed above, task analysis methods differ depending on the aspects of the task which they focus on (e.g., behaviours, knowledge, reasoning) and this leads to differences in how the results are presented and the type of recommendations that result (Adams et al., 2013). Therefore, before selecting a method of task analysis, the task to be analysed must be chosen so that the analysis approach suits the task being analysed. Resnick (1975) advises that an important consideration is the instructional relevance of the task and that the one selected should be decomposed because it is a task that will be taught, not simply because it is a 'low hanging fruit' for study or analysis. In this context, the task must be relevant to a dental operative curriculum, however, further criteria must be applied to ensure that the choice will not unduly prejudice future work:

- The task which is selected for decomposition will be compared to a traditional whole-task teaching approach using VR. Therefore, the operative procedure must be compatible with the Virteasy Dental Skills Trainers (HRV, Laval, France) available at the host institution.
- To ensure that differences in acquisition across the two instructional approaches can be explored, the task must have a significant cognitive component.
- The task should be measurable such that there are suitable discriminators at different levels of performance. Aspects of the indicators should also be identifiable in different contexts so that the transfer of the instruction can be measured.
- The task should also be understandable and not rely on existing knowledge of advanced dental operative procedures. This criteria contributes to several requirements: A complex task would reduce the pool of participants who would be eligible to take part; A task that relies upon existing knowledge could introduce expertise reversal effects that would bias the outcome; VR is predominantly used in preclinical settings (Towers et al., 2019) and therefore decomposing a task relevant to this stage would be the most impactful; finally, a simpler task will likely be more readily decomposed within a reasonable time period by the experts who will contribute to the analysis as participants.

The operative task of caries removal and cavity preparation fulfils all of these criteria. The Virteasy Dental Trainer has a library of existing carious teeth exercises that can be used. A number of competing factors must be accommodated and decisions made in the execution of the task, fulfilling the criteria of a significant cognitive component. The concept of “having a filling” is commonly understood by laypeople so the participant pool is preserved to the greatest possible extent. Finally, it forms part of the basic operative skills curricula across Europe and is recognised and assessed as an essential skill for students to develop during undergraduate training (Field et al., 2020).

A prepared cavity requires different attributes depending on the material selected to restore the tooth, so for the purposes of this project the task analysis of the caries removal and cavity preparation will be carried out with the intention of placing an amalgam restoration. Whilst at the time of writing amalgam is being phased out for dental restorations across Europe (EUR-Lex, 2017), the cavity design when using this material has additional attributes that can be used as differentiators of performance. In testing the effectiveness of the part-task/deconstructed approach this provides a greater opportunity to discriminate between performances. Cavities designed for amalgam restorations still share many design features with those intended for other materials so decomposing the amalgam cavity preparation task gives the greatest opportunity to validate the approach but still provides a subset of useful exercises after the material is phased out.

Having selected this task the difference between task analysis and functional analysis should be acknowledged. The focus here is on the decomposition of the operative task itself. A functional analysis would include broader considerations such as taking a history from the patient, identification of the presence of caries, acquisition of diagnostic radiographs and so on. A practitioner might expect these wider considerations as part of the task, but they are not currently simulated on the devices used for the study. Therefore, for the purpose of this project these preceding steps should form part of the briefing material such that the subject matter experts are aware of the assumptions and assist them in focusing on the task intended for decomposition. In many cases the description of the wider process is an indispensable precondition for successful task analysis as it provides the task analyst familiarity with the field and jargon. However, given the familiarity with the task of those involved in the project this step can be safely bypassed via the statement of assumptions (Piso, 1981).

The selected task is not without potential issues, it is possible that, despite careful consideration, the procedure results in too great a number of steps to be practically evaluated, or that the decomposed steps do not lead to tasks that can be taught using abstract models. Indeed, for breaking a complex task down into parts, a decision must be made on what a ‘part’ is (Annett et al., 1971). Some topics are awkward to teach and others fail to reveal the subject matter transparently (Resnick, 1975). Furthermore, in real life many tasks and the sub-goals are executed in a sequence, but time is not the only sequence and sometimes safety and emergencies must be attended to. Therefore, the priorities of the goals may be considered of greater importance than the order (Annett et al., 1971). Being aware of these risks is important at this stage as it may permit their accommodation during the decomposition process, but how they will be overcome will be considered during the instructional design phase based on the insight gained from the task analysis itself.

5.3 Task Analysis, Information Gathering and Knowledge Elicitation

There are many task analysis approaches and depending on the context, one of these formalisms may yield better results than another. These task analysis approaches are supported and enabled by a number of knowledge elicitation techniques used for gathering information. Having chosen an operative task for analysis, existing knowledge elicitation and task analysis methods were reviewed and evaluated for their appropriateness for the selected task. From this review, a novel evidence-based method was developed to analyse the task of interest in this work.

Task analysis methods provide a framework inside which knowledge elicitation techniques provide the tools for information gathering. A task analyst can draw on any of the techniques to meet the goals of the task analysis at hand but some techniques are more appropriate than others so their use must be guided by the selected task analysis framework. The knowledge elicitation techniques and task analysis methods reviewed for this project are detailed in Appendix B. Additionally, for each of the methods identified, a broad search of the literature in the English language was performed using Web Of Science™to explore if any of the approaches had been applied in a dental context previously. Unless otherwise noted the form of the search was:

Any Field: *Name of task analysis method* AND Any Field: Dent*

Results of this query are presented and discussed with each analysis method considered and reported in the aforementioned appendix. The next subsection will detail the selection of the overarching task analysis method used for this study.

5.3.1 Selection of the Overarching Task Analysis Method

Different requirements and educational objectives require different task analysis methods (Drury, 1983; Jonassen et al., 1999, p. 5) and the purpose of the selected method is to convert task-orientated data in to a representation useful for the project (Diaper, 2004, p. 34). Task analysis is the process of collating the information necessary to decide what to train (Annett et al., 1971), so it is important that consideration is first given to the goals that the analysis is expected to achieve. The description resulting from the analysis tends to assume the characteristics of the approach used so the instructional goals of the project should inform the selection (Jonassen et al., 1999, p. 6). Usually, a task analysis would be expected to have been carried out prior to the creation of the simulator as this would permit all human factors to be considered at the time where they are most readily accommodated (Kirwan and Ainsworth, 1992, p. 17). However, in this work, the simulator is already deployed and it is a specific application of the simulator that is being explored using task analysis, therefore, somewhat unusually, simulator itself must form part of the analysis.

Any of the approaches detailed in Appendix B could be used to produce a task description, so Annett et al. (1971) provides the following useful questions when evaluating an approach:

- Does the information obtained lead to training recommendations? Some methods record everything, which will include lots of redundant information making analysis more difficult.
- Does the method apply to more than a limited range of tasks? Some methods are more suitable for certain tasks than others, so it must be appropriate for the desired outcome.
- Does the task analysis method have any formal or theoretical justification? Whilst arguably not essential if the approach achieves the necessary outcome, a well tested and justified method will likely be more predictable.

These are all valid questions, but to support this work the following additional factors must also be taken into consideration:

- Subject matter experts will be participating on a voluntary basis. Some methods require a significant time commitment which may be unacceptable to volunteers, so the method chosen must be balance the time commitment with the fidelity of the analysis.
- As noted above, the simulator itself forms part of the system so the analysis must also explore where the simulator is deficient when compared with the real-world procedure so that these limitations do not impact on the developed exercises.
- The method must reveal the cognitive aspects and dependant knowledge used in the execution of the task. This is to permit the exercises developed from the analysis to reveal the underlying structure of the subject matter, be easy to demonstrate or teach, provide the trainee with a suitable plan of action (Annett et al., 1971) and ultimately allow understanding to be demonstrated by the learner (Resnick, 1975).

The review reveals that Task Analysis methods have seen very limited use in a dental context to produce learning materials. Whilst inspiration can be taken from the approaches used, the unique requirement of including the simulator as part of the analysis means there is no directly applicable method. Therefore the goals for this task analysis will be stated and the best-fit approach selected.

The goals of the task analysis for this project are to produce:

- A hierarchical break down of the steps involved in the procedure
- Insight in to the decision making used during each step
- A list of goal attributes that the successfully completed procedure should exhibit
- Common attributes that a poorly executed procedure would exhibit
- Discriminators that can be used to differentiate between levels of performance
- A robust description that is broadly agreed as the way that the procedure should be carried out at the host institution.

The description arising from this step will be used to develop a series of exercises to train the cognitive aspects of the task. However, because these exercises will be deployed in a VR simulation environment (rather than the real-world) they must incorporate the limitations of the modality to lead to the most effective educational use of the system. Therefore, it is an additional goal of this task analysis to identify areas where the VR simulation deviates from what would be encountered when carrying out the procedure in the real world.

Of the task analysis methods reviewed, two approaches appear to be the most useful in order to achieve this project's goals: Cognitive Task Analysis (CTA) and Hierarchical Task Analysis (HTA), which are amongst the most popular of the approaches to task analysis (Salmon et al., 2010). Given the cognitive focus of this project CTA is on first impressions a strong choice. However, the Cognitive Task Analysis family of approaches are suited to analysis of cognitive tasks, rather than the cognitive aspects of overt tasks. This subtle but crucial difference is relevant because, fundamentally, the task being analysed is a psychomotor task and whilst the cognitive aspects are a primary area of interest, neglecting the motor aspects will likely lead to a model that fails to describe the observable part of the task. These observable aspects are what determine the success of the procedure and what are largely used to measure the performance of a student. Furthermore, a CTA approach represents the environment in which problem-solving takes place as a mental model where learners construct their interpretations of the system they are working on and the processes required to manipulate that cognitive representation (Jonassen et al., 1999, p. 108). In a psychomotor task, the system state is not exclusively stored cognitively, it exists in the physical world so neglecting this co-dependent representation of system state does not appear to adequately encapsulate the problem domain.

The other approach that appears to be most suitable is HTA. HTA is more of a philosophy than a strictly prescribed approach that can incorporate a number of techniques(Annett, 2004, p. 402). When completed, an HTA produces a detailed description of the task in terms of a hierarchy of goals, sub-goals, operations and plans (French et al., 2019). The fundamental unit of analysis, and one of the defining characteristics of HTA, is the Goal. These are organised in to a hierarchy and further subdivided into subgoals and objectives (Annett, 2004, p. 403). HTA is a simple approach that collects data about a task through observation, questionnaires, interviews, walkthroughs and documentation review and was developed to better understand complex cognitive tasks in terms of what an operator is required to do to achieve the goal (Salmon et al., 2010).

This goal-driven structure should facilitate a description of the task that allows each goal or collection of complementary goals to be readily converted in to an exercise. Furthermore, HTA is more appropriate when analysing an existing system in order to make incremental improvements and modifications (Salmon et al., 2010). Given that this project is analysing the utility of a pre-existing simulator system in order to improve their utility with a series of new exercises, HTA can be argued to be a more appropriate choice.

Finally, a key area of investigation for this project is to decompose the task with a detailed description of each step. Whilst Annett (2004) suggests that Task Decomposition is already part of HTA, Kirwan and Ainsworth (1992) identify it as a distinct approach. Regardless of this distinction HTA can be integrated with other analyses (French et al., 2019), therefore, this project will adopt a hierarchical task analysis approach supported by a simultaneous Task Decomposition of each step. The Task Decomposition will be used to focus attention during the analysis on areas of specific interest and produce a robust description of the cognitive aspects of the task.

5.4 The development of a Task Analysis method using Virtual Reality Simulation

Having determined the overarching task analysis approach to be adopted, it is necessary to develop a method through which the goals of the analysis can be achieved. The flexibility to tailor an approach to a specific situation is a strength of HTA (Salmon et al., 2010) but with reference to the noted limitations above and in Appendix B the choices made require justification. This section describes the method in detail and states the theoretical basis for each component of the designed approach.

5.4.1 Overview of approach

Before describing the theoretical underpinnings of the intended approach, it will be helpful to provide an overview. All participants who take part in the task analysis will be qualified dental practitioners who's scope of practice includes the operative resection of tooth tissue and who are familiar with the Virteasy simulators available at the host institution:

- Participants will attend an individual simulation interview session where, in the presence of the investigator, they will:
  - Provide an initial list of the goals and subgoals in the task of caries removal and cavity preparation
  - Carry out each of these goals using a simulated carious tooth exercise on a Virteasy Simulator
  - After completing each goal, participants will pause to answer Task Decomposition follow up questions and provide clarifications
- After the individual session the notes and recording of the session will be transcribed into an individual task description. This will be shared with the participant for their own records and to optionally provide comment.
- When all participants have completed the individual session, the individual task descriptions will be combined into a draft of a consensus task description. Any ambiguities or areas where participants disagree will be noted for discussion.
- All participants will attend an online group consensus meeting where the draft will be presented, notes for discussion resolved and a consensus task description agreed.

The remainder of this section will take each of these steps in turn and provide a referenced description of the approach adopte, beginning with the use of Hierarchical Task Analysis and Task Decomposition.

5.4.2 Hierarchical Task Analysis and Task Decomposition

Hierarchical Task Analysis (HTA) with Task Decomposition will be used to capture a task description of a caries removal task in a Lower Left 6 molar and preparation of the cavity for restoration with amalgam. These approaches with the 'think-aloud' capture technique are regarded as the most relevant when performing a task analysis with the intention of designing training (Kirwan and Ainsworth, 1992, p. 22). HTA prompts the analyst to re-describe a task in to a hierarchy of sub-tasks and operations necessary to achieve a task's goals and the conditions under which they must be carried out (Kirwan and Ainsworth, 1992, p. 104). It is an iterative process that involves the gathering, preparation, organisation and analysis of data in order to recommend solutions based on the findings (Adams et al., 2013). This approach will be supplemented with the use of TD which is a structured approach to expanding the detail from a task analysis to capture statements from subject matter experts about areas of particular interest to the analyst (Miller, 1953).

The starting point for Task Decomposition is often a set of task elements derived from a pre-existing hierarchical task analysis. From this basis, the analyst can list these elements on an information sheet along with prompts for the additional categories of information that are required (Kirwan and Ainsworth, 1992, p. 97). To create these prompts, Miller (1953) suggests that each task element can be decomposed in to 9 categories and Piso (1981) suggests 8 questions that should be explored for each task element. However, it is recognised their suggestions might not cover all issues of interest to the analyst therefore it is often necessary to propose other decomposition categories (Kirwan and Ainsworth, 1992, p. 97). It is recommended to decide what information is required before recording (Kirwan and Ainsworth, 1992) so this study will perform a detailed Task Decomposition with a particular focus on factors around the VR implementation of the operative procedure. The task decomposition category prompts used for this study were created following the guidance in Miller (1953) and Piso (1981) but written to ensure that the areas of interest for this particular work were fully explored. Furthermore, because the format of the task analysis itself is not a concern (Kirwan and Ainsworth, 1992, p. 96) the decomposition categories will be collected concurrently with the task description itself during the simulation sessions for later organisation.

Where relevant to the particular step, participants were asked to describe in detail the following Task Decomposition areas of interest to this study:

1. Attributes of performance; how to identify when a particular step has been carried out acceptably

2. Common mistakes; as would be made by a learner when acquiring the skills to carry out the task

3. VR Simulation differences; identification of where the current realisation in the VR simulation context differs from what would be encountered in the real world procedure and how were these differences accommodated.

4. Equipment used; which tools were selected for each step and why

5. Attentional focus; where was attention focussed and what was being looked out for whilst executing the step

6. What cues revealed that a step was complete and the subsequent step could be undertaken

These categories of information provide insight in to the current limitations of the simulation environment which can guide the development of new exercises by recognising the factors that are being considered by the operator in the course of carrying out the procedure. Task Decomposition can require a significant amount of time, but it ensures that by systematically considering the areas of interest the resultant Task Description includes comprehensive details of the task and captures the information required for its later use (Kirwan and Ainsworth, 1992, p. 104). Doing this concurrently with the Task Description minimises the amount of the subject matter expert's time required and by planning ahead to consider the purpose of the analysis, only the most beneficial information is captured and no time is wasted capturing information that does not help achieve the analysis' goals (Kirwan and Ainsworth, 1992, p. 78).

Think-aloud protocol

Analysing complex tasks is best done with the collaboration of a subject expert (Kirwan and Ainsworth, 1992, p. 113), consequently, the first step of the task analysis is to gather independent descriptions of the task from each participant. These descriptions will be captured using a think-aloud verbal protocol of the performance using a VR simulation of the task with the researcher assuming the role of task analyst.

A verbal protocol gives accurate and (when needed) detailed information about motor performances and decision making using an efficient capture method that takes little additional time than is required to perform the task itself and in favourable situations only imposes a negligible cognitive load (Ericsson and Simon, 1980). Verbal protocols can be assumed to provide a basis for investigating the underlying cognitive processes and have a high face-validity with participants meaning that the technique is credible and understandable to non-specialist participants (Kirwan and Ainsworth, 1992, p. 79).

However, verbal protocols are limited by the fact that narrating a performance interferes with the performance itself, changing its speed or method of execution (Kirwan and Ainsworth, 1992, p. 71). Verbal protocols can be particularly challenging for describing motor performance because language may be inadequate for describing hard-to-verbalise tactile interactions (Steinberg et al., 2007). In the task used for this study, this would be a factor because it is a 'level 2' verbalisation task where the internal representation of the task is not in held in verbal form so has to be translated in to it. This translation is not thought to change the information from the way it is attended to (Ericsson and Simon, 1980) but the verbalisation is often concise and incorporates many personal idiosyncrasies. When the verbalisation is needed to communicate the information to another person, additional processing is required to find understandable referents that would be shared by the listener (Ericsson and Simon, 1980). Furthermore, the verbal reports can reflect the results of the cognitive process rather than the processes themselves because various intermediate processes may intervene between the internal representation of the information and its verbalisation (Ericsson and Simon, 1980). This can require additional probing on the part of the task analyst and can still be biased or inaccurate by the ability of the participant in narrating their performance (Kirwan and Ainsworth, 1992, p. 79).

To overcome these limitations, a number of approaches are recommended, which were incorporated in to the study design:

1. It is recommended that information is corroborated from more than one source. Individuals differ in their abilities to verbalise mental processes (Kirwan and Ainsworth, 1992, p. 79) therefore multiple participants will be recruited so that any deficiencies in one task description can be compensated by another. These multiple reports will be combined to produce a single description and then further corroborated via a group discussion of the combined description.

2. The quality of the description does not necessarily reflect the quality of the task performance so no preference was given in participant selection to those with a minimum number of years of experience (Kirwan and Ainsworth, 1992, p. 79). The individual verbal reports will be also cross referenced with a video recording of the performance and a documentation review to produce a written individual Task Description which will be shared with the individual participant to provide any comments.

3. Verbal reports such as a think-aloud process can produce around 150-200 words per minute so it is not practical to capture the performance as text. Therefore, it is essential to make a recording of every utterance (Kirwan and Ainsworth, 1992, p. 74). Additionally, a video recording permits cross-checking of the consistency of the report with other behaviour (Ericsson and Simon, 1980) because it is unlikely that the narrative will provide all of the information required (Kirwan and Ainsworth, 1992, p. 74) particularly given that subjects tend to stop verbalising altogether when working under heavy cognitive load (Walker and von Bergmann, 2015).

4. The think-aloud protocol was selected because it is rarely sufficient to rely upon observation alone as the reasoning behind decisions is unlikely to be observable (Kirwan and Ainsworth, 1992, p. 114). However, not all decisions will be verbalised and interrupting for additional detail will tend to disrupt the performance, but asking at the end of the performance may mean that the information that was in mind driving a decision could have faded (Kirwan and Ainsworth, 1992, p. 114). Therefore, to combat this effect, participants were asked for their 'outline' definition of the task's steps at the outset of the performance and then to pause their description at the boundary of each step to allow for questions to be asked.

This approach balances the desire to not interrupt the participant's performance with the importance of capturing contemporary data about the drivers behind the actions taken.

The use of VR simulation in the task analysis procedure

The task analysis will be performed using a VR simulation of the operative task with a Virteasy Dental Skills Trainer (HRV, Laval, France). The use of simulation is well established in task analysis (Kirwan and Ainsworth, 1992) as a mechanism for analysis when the nature of the task does not permit direct observation of the task itself in a real world context. However, the use of simulation in a task analysis context can also confound the process because the task itself is artificial and may not contain all the features of the real task (Kirwan and Ainsworth, 1992, p. 151). However, it would be unethical to perform a task analysis using a real patient as it would divert the surgeons attention to providing a narrative when their attention should be given fully to the successful completion of the procedure. A retrospective protocol is often suggested as an alternative where the description is based on a recording of the real performance, however, this can introduce memory lapses and bias (Jonassen et al., 1999, p. 260) where what is reported is not what was actually what was passing through working memory at the time of the step execution. Only verbalisations that rapidly follow a thought process can be taken as reflective of what was actually in mind (Charters, 2003) and given the focus on documenting the cognitive process this is not considered suitable. Furthermore, recording a real patient would require further consent, approval from NHS research ethics boards and would be difficult to capture high quality images of what the operator was viewing. Finally, it is an area of interest to the task decomposition in this project to explore areas where the VR simulation and the real procedure differ so carrying this out in the VR context provides an opportunity to capture the task performance in the context of interest.

Previous uses of task analysis in the literature have asked the practitioner to describe the task from memory using a questionnaire followed by a semi-structured interview (Tjiam et al., 2012), arguing that a semi-structured interview is superior to a video or stimulated recall approach (and by extension, simulation) because experts are recalling from experience whereas these approaches lead to a description of an idiosyncratic procedure. However, this approach asks the performer to generalise from previous experience and encourages (if not in fact requires) them to speculate about their process (Ericsson and Simon, 1980). Participants producing a description from memory are likely to be unaware that the data they are providing has inherent limitations and when asked to do so adopt one of 4 approaches (from Ericsson and Simon (1980)):

1. They are aware of the general approach and can recall the steps and report them directly without referencing the specific behaviour;

2. they are aware of some parts of the process that they used during a specific example of carrying out the procedure and attempt to generalise that into the description that they report;
3. they remember a specific example of the procedure and report it in general terms to infer a generalised procedure;
4. or, they draw upon a variety of information, general knowledge and experience of how one ought to perform the task and report that as the description

From this, it can be argued that is no logical need to perform the task at all, however, it provides considerable help to the analyst (Drury, 1983) and leads to more complete and accurate overview of the task because supplemental observations can be made during the operator's performance (Miller, 1953). This proposed approach is viewed as the one which will be most effective at capturing the information required from the expert participants. It minimises the need to construct descriptions from memory and provides a controlled environment in which to capture their real world experience.

The use of VR as the simulation modality for Task Analysis process has not been explored in the literature so far. This project has an aim of producing improved exercises for the VR simulators, so incorporating the VR at this stage provides an excellent opportunity for the expert-participants to include their thoughts about the VR simulation in the context of the decomposition. This grounds the discussion in the modality and may reveal additional relevant insight into the nature of the exercises to be developed later. However, the use of this modality necessitates that participants be recruited from the pool of teaching staff who are familiar with the VR simulators so that they are able to offer this insight without the novelty of the simulation experience detracting from their task description.

Individual Simulation Interview

The task which is being analysed itself sits within the context of a wider operative situation. At the corresponding point this analysis begins in the real situation, the presence of decay has been detected through examination, confirmed with radiographs and has been discussed with the patient. When asking a subject expert to decompose the task, the absence of this wider context may mean they lack crucial information that would inform their approach to the task. In this study, this context will be accommodated by reading to the participant background and contextual information. This addition was in part inspired by the importance of clarifying the purpose of an activity system step when applying Activity Theory to Task Analysis by Jonassen et al. (1999, p. 165).

At the outset of the task analysis and decomposition session, participants will be asked to break the procedure down in to between three and six of the major steps/goals in the execution of the task. This 'gross analysis' step provides a broad overview of the task, allowing detailed analysis during the simulation activity to be performed more easily (Miller, 1953) and ensures that time is not wasted at the outset by prematurely delving in to the fine detail (Militello and Hutton, 1998).

Once the broad outline of the task has been produced, participants can proceed with carrying out the procedure whilst thinking-aloud pausing when they reach each of their previously identified goals. This process will be carried out in the presence of an investigator because participants can feel uncomfortable narrating their thoughts to a recording device, whereas when an observer is present it feels more like they are talking to someone and results in a higher quality description (Kirwan and Ainsworth, 1992, p. 77). For each step of the task, the participants will be asked to describe what they are doing and thinking whilst working as well as specific things they are looking for and responding to. Thinking aloud whilst working produces higher cognitive load and can interfere with verbalisation because the task itself can *crowd out* the verbal information from working memory (Charters, 2003). Furthermore, some participants require coaching on how to think-aloud but that can lead to a risk of "leading the witness" and further interfere with the thought processes underlying the performance (Charters, 2003). To mitigate these risks, the participants will be selected from the clinical teaching staff at the host institution. These participants will have a secure knowledge of the task and should find the process of thinking-aloud more natural because they are likely to do so regularly in their teaching role. Splitting the task along the lines of the major goals/steps identified by the participant further manages the cognitive load as it permits the end of each step to act as a 'breakpoint'.

Interrupting the participant whilst they are working would interrupt the natural flow of their description (Charters, 2003) so these break points also provide an opportunity to pause and expand on the information provided during the think-aloud segment. During these participant-identified points, the participants answer specific questions relating to the Task Decomposition and elaborate on any points of interest that were mentioned whilst thinking-aloud.

After each of the areas have been fully explored, it is important to conclude with an open-ended question to give the participant an opportunity to add any missing detail that is important for the analyst to be aware of (Jonassen et al., 1999, p. 257).

Individual Task Description

After the session is completed, the participant's description must be transformed and transcribed into a hierarchical and tabular transcription of what was described. The video recording permits cross-checking to ensure consistency of the verbal report with observed behaviour (Ericsson and Simon, 1980) because it is unlikely that the narrative will provide all of the information required (Kirwan and Ainsworth, 1992, p. 74) and subjects can stop verbalising when working under heavy cognitive load (Ericsson and Simon, 1980; Walker and von Bergmann, 2015). However, it is important that the analyst does not twist the words of the participant and good practice is to allow the participant an opportunity to ensure the transcription agrees with what the participant intended by their words, to correct any factual errors and add further thoughts or comments (Charters, 2003). In this study, this will be achieved by emailing the participant a copy of their individual task description to provide such comments.

5.4.3 Draft consensus Task Description

Once all participants have completed their individual task analysis sessions and have had opportunity to provide comment and corrections on their transcribed Task Descriptions, attention will be directed towards unifying these descriptions.

For many tasks, a verbal protocol is the most direct means of accessing the information about mental processes, however the information gathered should be corroborated to reduce the risk of individual bias (Kirwan and Ainsworth, 1992, p. 77) and increase the validity of the description (Messick, 1995). The names given to tasks and the way they are grouped is often a matter of convenience to the individual and identification can be fairly arbitrary, meaning that different sets of task names can emerge from an analysis which relate to the same thing (Miller, 1953). Additionally, every subject is unique and differing levels of expertise arising from different backgrounds and circumstances may also impact upon the descriptions provided and what they believe a novice needs to know (Militello and Hutton, 1998). Therefore, every participant should be regarded as a mini-case study (Charters, 2003) and any inconsistencies between should be returned to the experts in order to be resolved (Kirwan and Ainsworth, 1992, p. 103). Therefore, in order to produce a robust task description it is necessary to reach an agreement as to the description of the task between all participants so that the individually described elements are representative of the group's collective understanding of the task.

To facilitate reaching this agreement a 'first attempt' draft of a consensus task description will be compiled by the analyst drawing on all participant's individual descriptions. This was viewed as useful mechanism to save participant time by presenting a description that could be amended rather than attempting to create one from scratch in collaboration with all participants. Furthermore, it reduced the possibility of conflict between participants because the draft represents the analyst's understanding of the task rather than challenging any individual's description (Kirwan and Ainsworth, 1992, p. 103). Discrepancies or disagreements in the individual descriptions encountered whilst creating the draft consensus description could be noted for discussion and clarification during the meeting.

Unstructured group consensus discussion

To achieve consensus an unstructured group interview will be conducted featuring all participants. This is recognised as a simple way to gather expert participants together to reach a consensus that will result in a better description than could be created by any individual (Kirwan and Ainsworth, 1992, p. 157). By providing a verification step and the input of multiple participants, the validity of the description is further enhanced. The output of group interview should be a consensus of the task description drawn from all individual descriptions. However, there is a risk that by challenging the inconsistencies between the individual descriptions, an expert may feel that their statements are being challenged and not co-operate in the process. Therefore, it is recommended that the analyst provide their own understanding of the system and ask the experts to comment on its accuracy removing the conflict and risk that the experts feel compelled to defend their previous comments (Kirwan and Ainsworth, 1992, p. 103).

Following the group consensus discussion, the agreed amendments to the draft Task Description will be made and this document used in future phases of the project.

The approach described above will be applied to the Task Analysis of caries removal and amalgam cavity preparation. Both the effectiveness of the method itself and the resultant Task Description will be discussed in later sections.

5.5 Aim and Objectives

5.5.1 Aim

Produce a structured description of the steps/sub-steps and underlying cognitive processes used when carrying out caries removal and amalgam cavity preparation in a VR environment.

5.5.2 Objectives

- Gather individual subject matter expert perspectives on the steps involved in carrying out the procedure using a think-aloud protocol during a one-to-one VR simulation session
- Identify attributes of performance, common errors and underlying cognitive processes/knowledge used by an expert by applying a Task Decomposition
- Document areas of perceived divergence between the Virtual Reality simulation and the real operative situation and perspectives on the impact of these differences

- Collate individual perspectives into a group-consensus task description using an online focus group.
- Evaluate the Task Analysis approach developed for use in this study

5.6 Method

Participants will be subject matter experts, defined here as qualified dental professionals who's scope of practice includes the operative resection of tooth tissue and who hold a contract with The University of Sheffield for the delivery of dental operative instruction. Participants will be selected for their familiarity of teaching with the School's VR system so that any novelty of the simulation context does not interfere with the focus of this study.

The task analysis procedure will combine Hierarchical Task Analysis with Task Decomposition as a 3 step process. In summary:

1. Participants will individually describe the operative task in detail during a one-to-one simulation session and interview in the simulation suite
2. From the individual task descriptions the lead investigator will compile a first draft of a group-consensus task description where participants' descriptions are in agreement. Divergence in the descriptions between participants will be noted for discussion
3. Participants will be invited to an online focus group to discuss and modify this draft with a specific focus on areas where differences were noted in the individual descriptions. The decisions of the focus group will define the group-consensus task description that will be taken forward into future work.

5.6.1 Individual simulation session

Participants will be guided through the task analysis process to individually describe the steps involved and knowledge relied upon whilst carrying out the operative procedure (using VR simulation). The session will be video recorded so that attention can be focussed on the analysis rather than taking notes. Additional information is available in the Appendix C containing the simulation session schedule.

1. Background and context to the scenario will be read to the participant (See Scenario Background in Appendix C)

2. Participants will be asked to list what they consider to be the main steps (approximately 3 to 6) when carrying out caries removal and cavity preparation. These will be written down and affixed to the simulator in view of the participant

3. Using a caries removal exercise from the library on the VR simulator, participants will be asked to carry out each of the steps they have just identified. For each step they will be asked to:

   (a) Describe the goal(s) of the step

   (b) To carry out the step using the simulator whilst 'thinking aloud' to describe what they are doing and why

   (c) Indicate when the step is complete and then answer a series of task decomposition questions to elaborate on or describe aspects not covered during think-aloud (See Appendix C)

4. When all steps have been described, participants will be asked a closing question to see if any other relevant information should be added that was not included during the individual steps

5. After the completion of the session, the recording and participants' answers will be used to compile their individual task description as a Hierarchical Task Diagram and Task Decomposition of the steps. Participants will be emailed a copy of the documents produced from their input for their own use and to optionally provide corrections.

Management of the individual simulation session

Individual simulation sessions are anticipated to take approximately 45 minutes and will be conducted one-to-one in the VR Simulation Suite at the School of Clinical Dentistry by the lead researcher at a mutually convenient time. The session will be recorded (audio and visual) on a secured and encrypted recording device with the camera pointing at the simulator display in order to capture the interactions and discussion. This recording will be transferred on to a University storage device approved for research as soon as possible after the session. The recording will be used, as described above, to write up the individual participant's task description and will be destroyed at the conclusion of the research project.

The Task Analysis Form (See Appendix C) will not be used by the participant. It acts as a reference for the lead investigator, listing the areas of interest to this study including sample questions relevant to each area. The primary means of capturing information is the audio/visual recording, with the form and sample questions providing a means of ensuring no areas of interest are missed and to note any questions for further discussion at the end of each step.

5.6.2 Group-consensus task description and focus group

Due to individual differences, it is unlikely that any two experts will produce exactly the same task decomposition. For example, what is considered a single step by one expert may be considered two smaller tasks by another, and some may not consider it a step in its own right at all. Therefore, multiple perspectives are required to produce a more robust description of the task. These different descriptions will be brought together to produce an agreed group-consensus task description.

After the final participant has conducted their individual task analysis the analyses from all participants will be compared and, where in agreement, combined by the lead researcher to produce a draft of a group-consensus task description. Simple differences (e.g. phrasing or equivalent terminology) will be confirmed by clinical members of the research team, however more significant differences in the descriptions of the steps or sub-steps will be drawn out for further discussion by the participants during an online focus group. Additional information is available in Appendix C which contains the group consensus schedule, but in summary:

1. All participants will be invited to an online discussion at a mutually convenient time where the draft consensus task description will be discussed.
2. After establishing meeting ground-rules, the goal of the session will be described and the draft group-consensus presented.
3. Where differences between individual task descriptions were noted, participants will be asked to discuss them and arrive at a decision as to the optimal description of each step.
4. After all differences have been discussed, participants will be asked if there is anything they feel is missing from the task description.
5. As soon as possible after (or where practical during) the meeting the draft group-consensus task analysis will be updated with the amendments agreed in the focus group.

Management of the group-consensus focus group

Participants will be invited to the online focus group using a private event in their University Calendar at a mutually convenient time. The meeting will make use of Google Meet, provided by University IT Services.

The focus group is anticipated to last 1-1.5 hours, but this cannot be confirmed until the degree of divergence between individual descriptions is known. In the invitation email, participants will be informed of the expected meeting duration. The email will also inform the participants that whilst they may recognise elements of their own task description when discussing the group-consensus draft, no other participants know the source of any content. The invitation will remind participants that they have the option of withdrawing from the study.

At the beginning of the meeting, when all members have joined and explicit notification has been given by the lead investigator, the focus group will be recorded (audio/visual) using the built in meeting recording facilities of Google Meet. The recording will be stored on a University storage device approved for research use and all copies will be destroyed at the conclusion of the project.

Comments and discussions during the focus group discussion cannot be anonymised between participants and all participants will know each other from their respective roles at the School of Clinical Dentistry. Therefore ground rules, including being bound by the Chatham House Rule, will be established at the beginning of the meeting with participants being asked to indicate their agreement. However, the procedure under consideration is well understood by all participants and is unlikely to be contentious. It is hoped that individual descriptions will be broadly aligned, leading to an interesting, collegial discussion that would not materially differ from that which might be had during a staff meeting to standardise teaching approaches. If a consensus for any step cannot be reached, a preference towards over-description will be chosen with a note to state that a step was not unanimously agreed.

5.7 Results

Four participants were recruited and successfully completed all parts of this study. All participants were employed as clinical tutors at The School of Clinical Dentistry, The University of Sheffield and were experienced at teaching in the School's VR simulation suite. Three participants were registered with the GDC as Dentists and one was additionally on a Specialist list. Participants had an average of approximately 21 years post-qualification experience and an average of 11.5 years experience teaching at undergraduate level [1].

The individual task analysis sessions in the simulation suite had durations of approximately: 1h 20m, 40m, 50m and 30m for Participants 1-4 respectively. The recordings of these sessions were used to produce the individual task descriptions and resulted in task descriptions of 3490, 2270, 2010 and 1607 words in length respectively (mean= 2344). At the beginning of the individual simulation session, Participants' described the task as consisting of 5, 6, 5 and 4 sub-goals respectively. When analysed in detail from the

[1]Due to the small number of participants and the size of the eligible research population, the individual/ranges of experience and specific specialism are not stated here to preserve participant anonymity

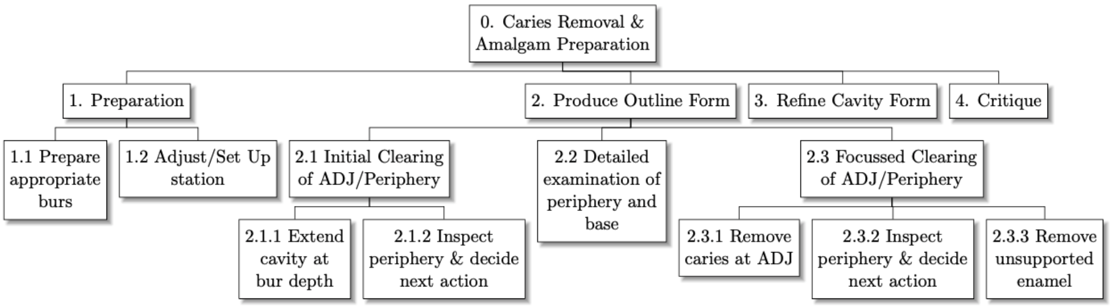


Figure 5.1: Participant 1, Individual Task Description

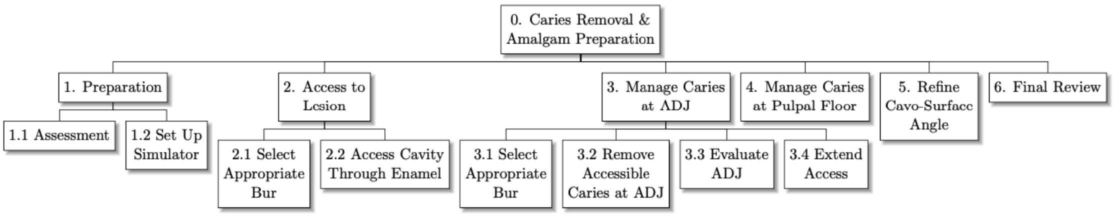


Figure 5.2: Participant 2, Individual Task Description

recordings, these resulted in to Participants 1-3 describing the task in terms of 14 sub goals and Participant 4 describing 12 sub-goals; differences of 2.8×, $2.\dot{3}\times$, 2.8× and 3× respectively.

Participants were able to view, and optionally provide feedback, on the write up of their individual task description. Two participants provided corrections and additional detail, one wrote to confirm that it was a fair reflection of their description and the final participant viewed the document but did not provide further feedback. The individual task descriptions are presented as simplified graphical overviews in Figures 5.1, 5.2, 5.3 and 5.4 and the full task descriptions (along with larger reproductions of the figures) are listed in Appendix D

After completing all individual task analysis sessions, a first draft of the group consensus task description was prepared for discussion at the group consensus meeting. This meeting was held online using Google Chat (Alphabet Inc, California, USA) where the draft was presented, discussed and corrections or changes agreed upon. The online meeting was of a duration of 1hr 15m in total and resulted in a consensus task description 7641 words in length with 21 sub-goals. A simplified graphical overview of the consensus task description is presented in Figure 5.5 and again in a larger format with the full text of the consensus task description in Appendix D.

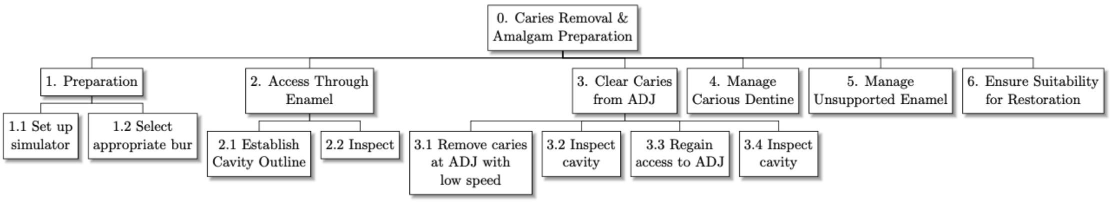


Figure 5.3: Participant 3, Individual Task Description

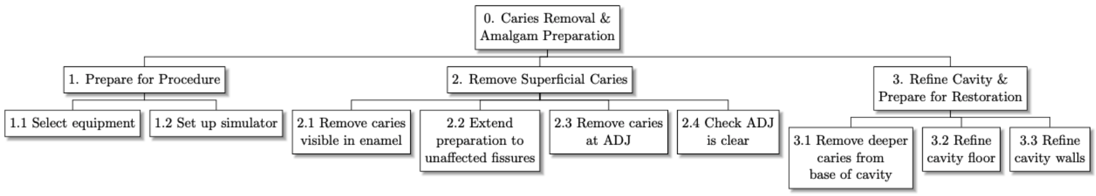


Figure 5.4: Participant 4, Individual Task Description

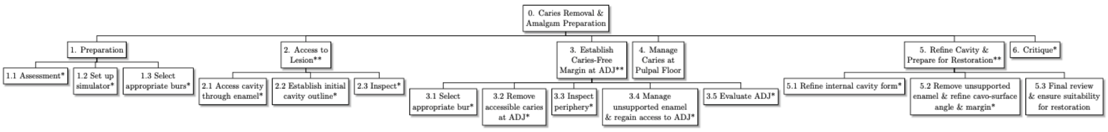


Figure 5.5: Consensus Task Description

5.8 Discussion

This section discusses the effectiveness of the method adopted and lessons from conducting the task analysis itself. The resultant Task Description and its implications for the use of Virtual Reality in pre-clinical dental education are discussed in Chapter 6.

5.8.1 Participant Selection

This study was reviewed and approved by the School of Clinical Dentistry, The University of Sheffield, Ethics Committee (Ethics Reference: 045165) and 4 participants were recruited to take part in the Task Analysis sessions conducted across the two weeks commencing the 25th April 2022. Between 3 and 5 expert participants is generally the recommended target as this allows gaining appropriate coverage of the task, balancing absences in the descriptions from one expert by those provided by another, but minimising the amount of repetition (Militello and Hutton, 1998; Charters, 2003). Analysis of verbal data is very time consuming so it is important to constrain how many interviews will need further processing (Kirwan and Ainsworth, 1992, p. 80) therefore, recruiting the median value of 4 participants could be viewed as ideal.

The method called for participants who were familiar with the VR environment which proved to have mixed results. In their descriptions participants often wanted to 'fix' the simulation and provided anecdotes from their experiences of teaching in this context. This deviation from documenting the task under analysis agrees with Kirwan and Ainsworth (1992, p. 72), who states that verbal protocols often lead to an operator pinpointing difficulties that they have with an existing system. There was also some suggestion that the tutors were already accommodating the differences of performing the task in a VR context and this impacted upon the approach they adopted and its consequent description. For example, Participant 1 can be seen in the video recordings to be using a 'backwards' hand position that would not be achievable in the mouth of a real patient and all participants were observed to use the view rotation controls rather than using the mirror. Perhaps domain experts who were unfamiliar with VR would have applied their real-world expectations and then commented more readily on the differences between the real world and simulation equivalent. But equally, the novice VR users may have been distracted by these differences and not been able to complete the task description in the allocated time. Experts have been successfully used to identify training deficiencies and to suggest alternative approaches (Kirwan and Ainsworth, 1992, p. 72) so there is likely merit in using both categories of participant. Future studies could evaluate if a different task descriptions are produced by participants with different familiarity in the simulation environment.

This study was perhaps limited by the fact that it only represents the collected opinions of tutors at a single institution and it could be argued that a more robust description would be created using a wider base of participants. However, given that the participants recruited described a range of approaches (see below), extending the recruitment to other institutions may result in benefits that are not proportionate to the increase of complexity of recruiting across multiple centres. Perhaps if the subject under analysis relied upon specialist knowledge held by only a small number of individuals scattered across different institutions then this may be warranted, however, here the participants were describing a well understood procedure from a Dentist's scope of practice. Furthermore, all task analysis should take into consideration the purpose for which the analysis will be used in order to produce a representation which is useful to the project's ultimate aims (Diaper, 2004, p. 34). Here, as long as the description provides sufficient detail to design and produce the exercises, it is not necessary for the analysis to be a fine grained description that captures every eventuality. However, should this be required in the future the description produced here can readily be expanded upon.

5.8.2 Individual simulation session

For the reasons discussed previously, the individual simulation session was captured (audio and visual) using two video cameras; one pointed at the simulator's screen to capture the procedure as carried out and the other framed to capture the participant's hands and their interactions with the simulator. Using a VR simulation rather than a phantom head negated the need for specialist intra-oral cameras as used in Walker and von Bergmann (2015) but similarly allowed further analysis at a later date and ensured no detail was missed. A video recording of the session meant that the analyst was free to focus on the decomposition aspects and identifying areas for follow up discussion rather than taking notes so this was considered effective.

The Standard Operating Procedures in force at the time due to the COVID-19 pandemic meant that at the time of the study the simulator's 3D glasses were not available for use. This meant that participants did not receive the full experience available from the simulators. Only Participant 1 noted the absence of the stereoscopic vision, feeling that it impaired their ability to judge depth and produce a finish closer to what would be possible in the real world. However, stereoscopic 3D is scheduled to be removed in future updates to the simulators so it is important that any future work does not rely upon a feature that will be soon be removed.

The individual sessions were anticipated to take 45 minutes. This duration was predicted empirically based on an approximate task duration plus an addition to account for the task performance speed being reduced whilst thinking aloud (Ericsson and Simon, 1980) and to allow for study administration such as consent. The median duration of the four sessions was 50m which suggests that this was a reasonable estimate however, during the sessions time was constantly being monitored to be mindful of participants other commitments so scheduling a longer session of closer to 60m would have permitted a more relaxed session and included time for further questioning to gather additional detail. However, the descriptions gathered were considered satisfactory and the fidelity of the description must be weighed against the additional time commitment asked of the participants and the ultimate purpose for which the task analysis will be deployed. Developing exercises for novices in future work does not require capturing every permutation and eventuality and it is possible that including these additional details would merely overwhelm a novice. Furthermore, gathering multiple perspectives and then verifying the combined description via the group consensus discussion allows the individual commitment to be minimised and for cross-compensation between participants to mitigate if any areas were lacking individually.

The differences in the durations of the individual simulation sessions also agrees the view that individuals differ in their ability to verbalise their mental processes (Kirwan and Ainsworth, 1992, p.79) and some require support and encouragement in order to be able to think-aloud (Charters, 2003).

The gross task analysis conducted at the outset of the session proved to be highly effective. These participant-provided high level goals provided natural break points aligned to their own conceptualisation of the overall procedure. This permitted points of interest from their think-aloud narrative to be expanded upon and depth added concerning their thought processes without inconvenient interruption of the task(Charters, 2003). Furthermore, the difference between the number of stages identified by the participants during this initial gross analysis and the stages identified in the transcribed individual task description differed by as much as 3×. This captures an important distinction where the 'from memory' rendition of the task differed substantially from the 'demonstrated' and narrated performance. Whilst this, as suggested by Tjiam et al. (2012), would be an idiosyncratic task it did result in a much more detailed description which may have been as a result of the procedural knowledge being reified to be recounted declaratively.

The task decomposition form (See Appendix C) listed all of the Task Decomposition questions and these proved to be a useful guide whilst observing the participant. Some decomposition questions were naturally covered during the participants narration whereas others could be annotated for further discussion. These questions were useful to achieve the Task Analysis as intended but in retrospect a further question would have been useful: one querying common areas of student difficulty. As will be discussed later, the questions included enquiring as to common mistakes made by novices and the attributes that correspond with a competent performance, however, these differ to those areas where novices specifically encounter difficulties, especially where this relates to understanding.

As recommended in (Jonassen et al., 1999, p. 257) the task analysis session concluded with an open-ended question to give the participant an opportunity to add any missing detail. Most participants felt that they had covered the necessary information, but Participant 3 took the opportunity to lament that students progressing to clinics should be able to apply the knowledge acquired in their pre-clinical courses to tackle real-world cases similar to the one presented in the task analysis. However, in some cases they were not as prepared as they should be for doing 'something real'. This observation suggests that something within the current educational approach does not transfer to the real world to the extent hoped. From this description there are many possibilities: perhaps the specificity of the motor skills prevents their transfer; perhaps current teaching practices where cognitive and motor skills are trained simultaneously hinder the student from applying the abstract knowledge elsewhere or perhaps the students are simply overwhelmed by the gravity of the situation they find themselves in when faced with providing treatment to their first patients?

5.8.3 Constructing the individual hierarchical task descriptions

Whilst the exiting literature describes, in detail, the information collection approaches and the abstract conceptualisations that can be used to classify the elements of a hierarchical analysis, there is a lack of description of the actual mechanism that one can use to take the record of the task performance and construct the hierarchical model. This could be due to Jonassen et al. (1999, p. 4) observation that Task Analysis is an *art* that is dependent on the skill of the analyst, therefore the exact approach will differ from person to person. This section will discuss the approach adopted in this work. All individual task analyses and their summary graphical representations are available in Appendix D.

The main source of information used for the creation of the individual task analyses was the video recordings of the session. Of the two cameras capturing the performance, the wider angle provided the most information for observing the interactions as this captured the participant's movements and mostly-sufficient view of the simulator screens. The screen-focussed recording was used to supplement the wide angle view when the performance could not be readily observed from the wide angle view if, for example, the participant's hand was obscuring what was being described. As suggested by Kirwan and Ainsworth (1992, p.74), it is unlikely that the taking written notes would have been able to keep pace with the quantity of information provided so the full recording was invaluable for constructing the hierarchical description. It is highly unlikely that capturing a written description of what was done whilst trying to observe the motor-skills based interactions, listen to the description and identify any relevant questions and clarifications to support the task decomposition would have produced a reliable documentation of the task for later assembly into a hierarchical form.

The creation of the hierarchical task description was an iterative process centred around repeated viewings of the performance video capture and the progressive building of a hierarchical model of performance, adding detail or making modifications in subsequent viewings. The first attempt at this process tried to base the hierarchy on the participant-provided gross task analysis recorded on Post-it® notes at the beginning of the session and then augment this with further detail from the recordings. However, whilst these Post-it® overviews helped structure the participant's description, they proved ineffective as a source of the hierarchical model. This is predicted by the literature, because the initial breakdown of the task was derived from what is normally automated non-conscious knowledge that is used 'live' (Basque et al., 2014) so is incomplete and did not fully align with what the participant actually did in the performance. The differences in the number of number of steps identified at the outset of the session and the number of steps actually described during the performance ranged from 2.3× to 3×. It is also predicted in the literature that participants consider the task at different levels of abstraction and there is a close negative relation between the degree of practice and the awareness of the intermediate steps in the procedure (Ericsson and Simon, 1980). This observation is supported by the participant with the most experience representing the task with the fewest initial steps, however, this study does not provide enough data to support this observation meaningfully.

Hierarchical task analysis, as suggested by the name, produces a hierarchical model of goals and the tasks or operations required to accomplish them. More complicated goals can be subdivided in to a series of sub-goals which must be completed to achieve the super-ordinate goal. Within a goal, plans describe trigger conditions and any prerequisites that must be satisfied before moving on to the next goal. To identify the goals, the recording was reviewed with a focus on identifying verb phrases or outcome-focussed statements. For example, phrases such as those beginning "first we need to" or "once we are happy that" were good indicators that a goal was about to be described. A short phrase encapsulating the described goal was written down on a sheet of A3 paper and the recording resumed. When the next outcome statement was encountered it was evaluated for if it was part of the current super-ordinate goal or if it should be listed as sub-goal with a series of actions beneath it. Sibling-goals were written down adjacent to the preceding goal and sub-ordinate goals immediately below it. Over a series of viewings, the model was refined to produce satisfactory hierarchical description of the task. In all cases, what was determined to be a sub-goal and what was a task within the super-ordinate goal was based of the judgement of the analyst, but with influence from the participant's Post-it® gross task analysis. For example, a described action could meet the criteria to be represented as a sub-goal but only require a very small number of actions to complete. In these cases it was sometimes judged that representing these separately would harm the clarity of the overall task description. This approach is consistent with the recognition that there are many ways in which a super-ordinate task can be broken down and an analyst should be wary of introducing too many task elements as this often confuses rather than clarifies the description (Kirwan and Ainsworth, 1992, p. 114). To conclude the assembly of the hierarchical model, the participants the Post-it® descriptions were revisited one final time to see if the wording of the analyst-derived goals could be revised to bring them closer to the wording used by the participants at the outset of their task description.

It was readily apparent from the length of descriptions that an exclusively diagrammatic representation would be insufficient to record the outcome of the analysis. Whilst more easily assimilated, diagrams are harder to annotate with detailed design notes (Kirwan and Ainsworth, 1992, p. 112). Therefore, instead of annotating the diagram, the goals were recorded as section titles in a separate document containing the detailed sequence of steps below each heading based on the think-aloud narrative provided by the participant. During the documentation of these individual steps, gaps were evident in the verbal descriptions. It is recommended that the verbal narratives are corroborated with other data (Kirwan and Ainsworth, 1992, p. 71) so the video recordings of the actions taken were reviewed and, where necessary, a wider documentation review was undertaken to fully establish the statements used for describing the steps within a task. The need for this additional verification is consistent with the literature because the inner vocalisations of what is being done are not usually intended to convey information to a listener so can lack the detail needed to be understandable by others (Charters, 2003).

The construction of and cross-checking required to produce the individual task descriptions was a time consuming process and was found to be broadly consistent with the 4-8 hours of analysis per 1 hour of video recording predicted by Kirwan and Ainsworth (1992, p. 78). If this or a similar approach were adopted for future work, the necessary time commitment should be factored in to any planning. However, whilst time consuming, the structure that this approach provides ensured that the areas of interest were systematically considered (Kirwan and Ainsworth, 1992, p. 104) and produced a highly detailed description of the task from each participant, revealing much of the underlying cognitive structure.

Once the hierarchy of the task analysis and the component tasks had been established, the task decomposition steps were added to the document. Task Decomposition systematically gathers more information about the task through focussed questioning based on a prior task analysis (Kirwan and Ainsworth, 1992, p. 103). However, because the decomposition step was intended from the outset the intermediary transcript that these are usually based on was unnecessary and an additional data gathering session was not required. Therefore, to compile the task decomposition information it could be simply transcribed from the recording and appended as appropriate within the document.

There is a risk that an analyst may twist the words of a participant so, as recommended in Charters (2003), each completed Task Description was shared with the corresponding participant to provide an opportunity to ensure what was written agrees with what was intended by their words and add any further thoughts or comments. Whilst under no obligation to do so, three of the four participants provided feedback on the task description. This feedback either corrected word choices that had been used from the video recordings or rephrased goals to more accurately capture the intent. Given the purpose of this investigation was to create an agreed description, ensuring that the individuals were happy with their own contribution was logical and provided the most accurate input to the group consensus stage. For example, Participant 1 corrected "Create traditional outline form" to "Create outline form" and Participant 3 requested that "Extend cavity with high speed" was modified to "Remove grossly unsupported enamel to further inspect ADJ". In both cases, the original text accurately reflected what was said during the sessions, but in the first example "traditional" was regarded as having undesirable connotations to the outmoded GV Black cavity designs, and in the second example provided a more precise description of the goal.

Sharing the task descriptions with the participants did however provide some mixed results. Whilst it allowed desirable corrections such as the above to be made, the participant's lack of familiarity with the hierarchical notation (especially with looping constructs) led to some 'corrections' that were already catered for in the description. Sharing a task description for a participant to review in their own time is not generally recommended because it allows the reader to project their own understanding to what was written or simply to skim over what they assume to be an accurate transcription of what they said (Kirwan and Ainsworth, 1992, p. 115). Therefore, a further meeting with the individual participants to present task description is desirable. Additionally, a further meeting provides an opportunity to further quiz the participants on any areas that seemed clear during the individual session but raised questions during the transcription. However, the participants were under no obligation to provide this feedback so requesting a further time commitment to present the transcribed hierarchy and add further detail to what is in effect an intermediate step in the process may not be an effective use of time. This is especially true given that the gathering of multiple perspectives is intended to cancel-out any shortcomings in an individual description when the draft consensus document is compiled. Furthermore, participants were invited to provide feedback and discuss the draft of the consensus description at a later stage.

5.8.4 Comparing Individual Descriptions of the Task

Capturing these individual task performances provided an interesting insight in to the conceptualisation of a common task across four highly experienced subject mater experts. Whilst a full comparison of the operative approaches and their clinical significance is outside the scope of this work, some of these observations from both the individual within-subject performance of the task and a between-subjects comparison have relevance to the use of VR in the training of pre-clinical skills.

The first of these was a recognition that some ambiguity existed as to the boundary of the task used in this work. This ambiguity led some participants to describe activities that were considered to be outside of the specific task under consideration and part of a wider functional analysis. Naturally, some grey areas are to be expected: for example, bur and handpiece selection precede the performance of the task but fall within the decision making necessary to perform the first step. However, some participants took this further back and discussed the importance of performing a thorough assessment of the case including a visual inspection, a review of pre-operative radiographs and a review of the patient's notes to establish any history of caries. Participant 3 even described the importance of prior encounters with the patient and personality factors as influencing the approach, for example if the patient nervous or prone to 'wriggling in the chair' then that may guide what is achievable in the present situation.

The design of this study attempted to guide the participants towards the part of the procedure under analysis by reading a pre-task background. It is possible that this contextual information was not clear enough and led some participants to discuss a wider functional view of the task, however, had this ambiguity not existed an interesting observation would not have been captured. When providing the treatment in a real clinical context, all of these wider factors as volunteered by the participants would be taken into consideration in order to determine the most appropriate treatment and approach. However, these considerations are not currently part of the VR simulation experience and if importance is not given to them in a training context, then can it be expected for a novice student to give importance to them when the skill is transferred to the real environment? Participant 3 mentioned that 'is restoration indicated as necessary' is a question that every student should be asking themselves before carrying out a procedure, however, the very virtue of loading an exercise on a VR simulator frames operative intervention as a fait accompli. To what extent, especially amongst junior of students, does this create an assumption or form habits that when a patient is in the chair that they will necessarily require operative treatment? Perhaps the VR exercises should include an assessment stage and ask the learner to review relevant case information to confirm that a restoration is indicated and consider relevant risks of iatrogenic damage (e.g. pulpal exposure) before continuing with the task so that this vital step is given the rightful place in their decision making?

There were a number discrepancies between what was done and what was verbalised by the participants which introduced issues when transcribing the task descriptions. For example, most participants described their motions with the bur as "continuous" however, the video recordings showed a sweeping or repeated brushing motion. It is possible that the reasoning for the overt behaviour may not be in short term memory and the participant is inferring a description (Ericsson and Simon, 1980) perhaps defaulting to the same wording used during their own training. This was discussed during the group consensus stage, and it was agreed that a sweeping motion was closer to the overt behaviour however, a full investigation of this discrepancy is outside the scope of this work. As discussed above, it does serve to illustrate the importance of capturing the motions and not simply relying on the narrative in the creation of the task description.

Between the participants there were clear differences in their propensity to describe the task in terms of goals. Participant 1 described the process of caries removal as more of an iterative process where the cavity formed progressive *evolutions towards* and *approximations of* the desired form which did not lend themselves to a goal-orientated description. As the participant with the most experience, the task is likely to have become highly proceduralised and being asked to think in this way clearly appeared to be a challenge. Over the course of the individual session, the participant increased the number of steps from the 3 initial high level steps to 5, giving the impression that the structure of a well understood task was being reconstructed on-the-fly for conveyance in a verbal form. This is very much in line with the literature; much of our thought is not held in verbal form and a progressive abstraction of knowledge that develops with expertise means that words only form part of that representation's meaning (Charters, 2003). The automation means that the steps are carried out without being interpreted so the deeper reasoning is not available to short term memory (Ericsson and Simon, 1980) and only when the sentences are formed is the thinker is aware of it (Charters, 2003). This is in contrast to Participants 2 and 3 who, whilst acknowledging a degree of overlap between stages, very much represented the task in linear terms consisting of discrete steps with specific goals. Participants 2 and 3 mentioned working with more junior learners in their day-to-day teaching, so this might suggest that some of this structure was already available and could be drawn upon to describe the process, whereas Participant 1 was having to construct this representation 'live'.

In addition to the differences in task conceptualisation, Participant 1 also continued to use the high speed hand piece for longer than the other participants. Bernstein's stages of learning (Bernstein, 1996) suggests that at at stage 3, a performer is exploiting passive dynamics and has a level of skill associated with effectiveness and efficiency of motion. Participant 1 described that the progressive approximations of the final form required less swapping of the handpiece between high and low speed and this was quicker, more efficient and had less opportunity to introduce cross-infection. However, this requires planning ahead and a high degree of experience to visualise the end-form and work towards it. This would suggest that this approach is inappropriate for a novice. As their expertise grows, the novice would be able to audition similar movement economies and when they find their experience and finger dexterity is equal to the task they can take advantage of the time saving that they provide. That these differences of approach exist and were found here agrees with a limitation of task analysis that the structure and arrangement of knowledge for expert use differs from the optimal arrangement of knowledge for skill acquisition at any given stage of learner (Resnick, 1975) and reinforces the importance of conscious organisation of learning materials to provide the optimal learning approach (Cheng et al., 2015).

There were also more subtle differences noted between participant performances. The phrase "use the biggest one that fits" was a colloquialism used by Participants 2 and 3. However, when used by Participant 2 it was used in relation to the size of bur that should be used to clear the ADJ caries, but Participant 3 used it in relation to the clearing of caries at the pulpal floor. Clearly, to a novice learner the use of such phrases being used in subtly different ways risks introducing confusion so whilst these phrases are likely to be memorable and useful, care must be taken that there is a common understanding of their usage across teaching staff.

Participant 3 remarked that students struggle with clearing the ADJ and knowing what is required. This is notable because whilst there is broad alignment for the overall procedure, the 4 participants in this study all described the details slightly differently. Even if these 4 descriptions here represented the full extent of possible variance for this task, it could mean that a learner is receiving subtly different explanations from different tutors when seeking clarification on how to perform the task. This conflicting information would need to be reconciled by the learner with their existing knowledge creating cognitive dissonance (Harmon-Jones, 2012) and increasing the extrinsic cognitive load placed on them in the acquisition of the skill. It could be argued that the learner is perhaps not always simply struggling with a concept, but struggling to acknowledge that there is not a universally agreed approach to the task. This strongly supports the use of task analysis in the creation of learning content because it draws out the existence of these inconsistencies and encourages conscious discussion amongst educators to agree the way that the material should be structured. For this project, the use of a follow-up consensus discussion appears to be well judged to facilitate this discussion prior to the development of the future exercises.

5.8.5 Consensus Task Description

Once all participants had completed the individual session and had reviewed their individual task description, a draft consensus description, drawing on all individual descriptions, was produced by the analyst. This was then shared with the participants during an online meeting to allow comment and to resolve discrepancies encountered during its creation. The following sections describe the process through which the consensus task description was created.

Constructing the draft consensus task description

To create the 'first attempt' draft for the consensus task description the following process was followed:

1. Firstly, a high-level graphical representation of the goals taken from the individual descriptions was created and desk-tested for flow. Similar stages were merged and then individual goals were empirically re-described to produce a description that most closely represented the descriptions provided by the majority of participants.
2. Any descriptions of goals that disagreed with the majority for a given step were noted for discussion in a separate document.
3. The text of the goals in the merged representation were then copied as headings and subheadings (to represent the hierarchy) in to a new document.

4. The body of the relevant tasks and decomposition headings from each participant were then copied beneath the relevant merged heading and colour-coded to differentiate the individual sources.

5. The body-text for each of the consensus-goals were then then addressed one at a time, unifying to a single description of the goal drawing from all participant perspectives. Again, where un-resolvable differences were encountered, these were noted for discussion.

6. Merging the task descriptions required an iterative process to progressively amend the high-level definition of the goals and re-check their accuracy. The boundaries between tasks are not always clearly defined so it can be necessary to return to the descriptions to further break down the task (Kirwan and Ainsworth, 1992, p. 103). This manifested itself where individual descriptions did not suggest that a sub-goal was necessary in isolation but when combined with the other descriptions providing alternative detail, some tasks were found to warrant their own sub-goal in the merged consensus description.

The creation of a formal task description based on information from various sources is very much a creative endeavour relying on the intuition and skill of the task analyst (Jonassen et al., 1999, p. 4) so there are clear opportunities for bias and mistakes to be introduced at this stage. However, as discussed above, the existence of draft allows the expert participants to make modifications against an existing document rather than trying to reconcile the differences and build consensus from scratch during the meeting (which would necessitate a significantly longer meeting). The opportunity for the participants to review and confirm the draft mitigates these risks whilst keeping an acceptable demand on participants time .

The multiple perspectives gathered as input to this step proved effective here. Multiple attempts and differing descriptions of the task meant that, as predicted by the literature, deficiencies in the the individual descriptions were compensated for by another. For example, Participant 1 did not describe a discrete clearing of the pulpal floor stage because caries from the floor was satisfactorily cleared during earlier stages. The participant commented upon this when providing feedback on the description, however, because other participants did describe this step, it was properly included when preparing the consensus task description.

The draft consensus description reached 21 sub-goals, which is 1.5× more than the most identified by any individual, and the consensus description totalled 7641 words, almost 3.6× the average length of the individual descriptions. This suggests there were many areas where each subject matter expert attended to different aspects of the performance and provided differing amounts of detail for each step but the overlap of such detail was fairly small. Given the length of the merged description in comparison to the the individual descriptions it is possible that additional participants beyond the 3-5 recommended would describe the task in even more detail, however, there is likely to be a degree of diminishing returns which this study was not designed to investigate.

Group consensus meeting

After the completion of the draft consensus, an online meeting using Google Meet (Alphabet Inc, CA, USA) was scheduled and attended by all 4 participants. Consent was re-established and then the discussion was recorded (audio and visual) using the built in recording functionality of the Google Meet platform. After establishing ground rules for confidentiality, the draft consensus task description was presented and verbally walked through (as recommended by Kirwan and Ainsworth (1992, p. 114)). After completing the walk through, participants were asked for their initial reactions followed by specific discussion around questions that had arisen in the production of the draft description. The meeting itself was scheduled for 1 hour in duration, however, due to connectivity issues and the breadth of discussion participants were happy to extend the the final duration to 1h 20m. The complete task description as agreed at the group consensus meeting and a graphical summary of the process is presented in Appendix D

Participants' reactions to the draft was surprisingly positive. Despite drawing from 4 differing perspectives and resulting in a description that differed to any individual description of the task, it was universally accepted by all participants as a good, highly detailed and comprehensive description of the steps necessary to perform the task. Participant 2 commented that it was surprising how much is involved in the procedure when it is described in such detail and suggested that this contributes to why students struggle with acquiring the skill. A participant not recognising the complexity of a task they carry out frequently is very much in agreement with the motor skills literature discussed previously and shows that the automation has 'hidden' some of the complexity from their conscious awareness (Tjiam et al., 2012).

No major changes were proposed to bring the draft closer to any individual's description with the only modifications being simple clarifications and the resolving of ambiguities from the individual session recordings. Some steps that appeared to have very different conceptualisations were in fact quite similar when discussed; the differences simply arising from a different framing of the task. The most significant discussion was around the differences between what is 'really' done as an expert and how that differs from the 'by the book' approach that a novice should be taught and use. Participants agreed that experts develop a judgement as to which steps can be merged for expediency whereas novices should pause at each step-boundary to consider the state and plan their next actions. Again, this agrees with much of the previously discussed motor skills literature on the development of expert performance. This point was annotated against the relevant steps in the task description as a potential variance so that the insight was captured.

A broader discussion was also had around the observation that none of the participants picked up the simulator's mirror, instead opting to use the view rotation controls. During the individual session, Participant 1 commented in jest that the most convenient orientation was to access the tooth as if through the patient's larynx. This discussion introduced a number of points that are of relevance to the use of VR simulation in this educational context. Firstly, the tooth used in the exercise is mostly visible using direct vision in the real world so the use of the mirror is mainly for retraction of soft tissues rather than for vision. The main area where a mirror would be needed is when viewing the mesial aspect of the cavity, however, participants discussed that not forcing the use of the mirror could be considered an acceptable scaffolding for future learning. Depending on the learning objective of a particular session, the complications introduced by the mirror could be viewed as a distraction when the overall goal is to teach for example the clearing of all aspects of the ADJ. Interestingly, if the substrate being used for teaching has excessive fidelity for the learning objective, this supports the hypothesis that a reduced fidelity exercise may be suitable as it can avoid introducing these 'artificial scaffolding' conflicts between what should be done and what is needed for the particular learning objective.

5.8.6 Evaluation of the Task Analysis approach adopted

Performing a task analysis and decomposition in the context of this project served a number of goals:

1. To catalogue the steps taken by an expert in the performance of the task in such a way that the process was broken down in to logical goals
2. To capture insight into areas of particular interest in a task decomposition
3. To provide the source material from which to develop educational exercises that facilitate learning the cognitive skills necessary to achieve these goals
4. To do the above in the virtual reality context so that the limitations and differences between the simulation and the real task can be brought to the fore and can be accommodated in the design of the exercises.

The approach adopted drew from Hierarchical Task Analysis, supplemented with a Task Decomposition. These complementary approaches were felt to be the most suitable to achieve the goals of this project, the Hierarchical Task Analysis provided the overall structure for the investigation, and the Task Decomposition ensured that areas of interest were systematically considered (Kirwan and Ainsworth, 1992, p. 104).

Strictly, the task analysis carried out here was a partial task analysis. This work only conducted a single phase of analysis and considered only a single instance of occlusal caries on a single tooth. A more comprehensive task analysis would consider multiple cases and their individual variances and then extend the investigation to discuss caries on different surfaces of the tooth, carries affecting two or more surfaces of a tooth, caries occurring on different teeth and so on. However, this would result in a study producing far more data than is required for the goals of this project. The method adopted here could prove to be scalable and assist expanding to such a task description. An iterative approach may indeed prove to be preferable to attempting to capture every permutation in a single cycle of analysis as it would permit exploration of the model as it is elaborated upon. In this study, participants volunteered alternatives throughout their think-aloud process, discussing partial versus complete caries removal and how different sizes of cavity required modifications to the approach. Some of this insight into the decision making process would be valuable when developing future phases of exercises, however, for the purpose of establishing the utility of the approach proposed in this work it is only necessary to develop parts of the hierarchy where it is justified (Kirwan and Ainsworth, 1992, p. 117). Fundamentally, the wider goal of this project is to explore a deconstructed training model in a VR context and not to produce an internationally robust task description of caries removal and amalgam cavity preparation.

The point at which a task analysis is carried out in the wider project's lifecycle can have an impact on the nature of the analysis and what can be done with any insight produced. VR simulators have already been developed and released to market and it is somewhat unusual to deploy a task analysis approach at this stage. Task analysis is predominantly intended to be an up-front activity and should have been performed prior to the creation of the simulator as it would have allowed all of the human factors to be addressed at the time most suitable to accommodate them into the design (Kirwan and Ainsworth, 1992, p. 17). However, as VR simulation becomes more established and its use grows, there will be a natural desire to develop additional material for use with this modality, so an approach that captures a description of an operative task in the context of the limitations of the intended modality is a welcome addition to the literature and will aid the creation of more effective learning content.

However, it must be recognised that Task Analysis is an art and a skill that must be developed by the analyst through practice (Kirwan and Ainsworth, 1992, p. 118) and the results of which are dependent on the skill of the analyst (Jonassen et al., 1999, p. 4). Therefore, this does introduce reproducibility issues when used in a research context whereby the interpretative aspects mean that, even given the same inputs, two analysts would not necessarily produce exactly the same model. Perhaps, task analysis approaches in an educational context are best framed as an inductive approach to exploring the performance of a task rather than a deductive research approach. Three forms of research are needed to advance a field of science: inductive/exploratory in order to discover new areas for exploration, abductive/explanatory to establish feasible explanations and theories, and deductive/confirmatory research in order to test the validity of those theories (Woo et al., 2017). Here, the analysis has explored a conceptualisation of a task in a new context which has identified opportunities for further exploration.

The purpose for which an analysis will be deployed should be the guide to the level of analysis that should be carried out (Annett, 2004) so if the resultant task description is sufficient for its intended purpose then perhaps it can be argued that the lack of reproducibility is less of a concern. The process captured multiple viewpoints and merged them into an agreed description and whilst this does not contribute towards its reproducibility, it does contribute positively to the validity of the task description produced (Messick, 1995).

The use of VR in the task analysis of a dental operative task

The use of simulation is well established in task analysis for situations where the nature of the task does not permit direct observation in the real world context (Kirwan and Ainsworth, 1992, p. 151). A simulation based task analysis was performed to good effect in a dental context for a cognitive task analysis by Walker and von Bergmann (2015) to describe the steps involved in wax carving for a Class II restoration. However, in the present study VR simulation replaced physical simulation and resulted in an interesting task description that describes something between what is done using VR and what occurs when carrying out the procedure in the real world. The literature recognises that this can be a drawback of using any simulation based information capture technique (Kirwan and Ainsworth, 1992, p. 151) but it does appear very pronounced in the dental VR environment. Almost every participant identified differences at every step of the task and commented to their operative significance. This could reduce the effectiveness of using VR as a vehicle to capture a 'pure' task analysis of an operative task for use in a wider teaching context, however, this limitation is a strength for this project because these very differences were of interest and this insight will be invaluable to the design of the exercises in later work. The identified limitations are listed against each goal of the analysis in Appendix D.

The VR context's omission of various aspects of the operative environment and experience changed the way in which the participants interacted with the simulation. Participants identified that they modified and adapted their approach from what they would do in the real world execution of the task to accommodate or work around the deficiencies in the simulation. Some of these omissions are likely to be problematic in the use of the simulator as a teaching tool because it did not represent (in any way) crucial visual and tactile cues that are extensively relied upon operatively and which materially contribute towards the decision process. An expert can draw upon their experience and accommodate these limitations, focusing elsewhere to infer the state, but to a novice the current VR simulation is insufficient to prepare them for the real procedure by not representing many of the cues they should be attending to in the execution of the task. It is possible that practice in this environment may lead to unintended consequences and result in the learner focussing their limited information-processing capacity on the wrong stimuli (Schmidt and Lee, 2014, p. 40). Fundamentally, if important cues are not present in the simulation, how can the learner be expected to know they should be attending to them in the real world?

Evident in much of the above discussion is a fundamental question in any task analysis: was the task description produced sufficient for its intended purpose? In the context of this work this is likely to be the case. As discussed, there are many opportunities to add to or enhance the description itself and there are opportunities to enhance the process by which it was created. However, almost all of these would require additional time and effort from the subject matter experts. If the description provides the information necessary to design the instructional exercises then it is likely to be of sufficient detail for its purpose. The Task Analysis literature discusses that an analysis should stop when $P \times C$ approaches zero; that is when the cost (C) of failing at a sub-task multiplied by the probability (P) of a mistake occurring approaches zero (Mannan, 2005; Annett et al., 1971). In many cases this is a good maxim however, in this case the task is adequately described when the description is specified to the level where there is a shared understanding between those providing the description and those who will rely upon it. The use of VR to perform this analysis was suitable as the same deficiencies it would have in other contexts were a material consideration for the intended purpose of this analysis.

The above illustrates that using VR as a tool to capture the task description for an operative skill has limitations when attempting to capture a faithful description of the steps carried out in the real world. As such it is likely to be problematic where the intention is to provide training in a non-VR context. However, where the intention is to develop future exercises for a VR context, its use in the analysis is invaluable as it ensures that exercises developed for the modality accommodate the limitations of the modality. Accommodating these limitations permits the effective constructive alignment of the new exercises against the identified performance criteria (Jonassen et al., 1999, p. 1) and can prompt discussion on how to avoid or mitigate any unintended consequences of the training.

5.8.7 Conclusion

The use of VR simulation to decompose a dental operative task has not been explored in the literature previously. The task analysis performed here incorporated the VR simulators into the task analysis of an operative task that they are currently being used to teach. Task Decomposition was used to document key performance indicators and the expert-participants thoughts on the simulator's suitability for this task, focusing on exploring differences between the simulated and real world contexts. This grounds discussion of future interventions in the limitations of the modality and permits an exploration of the current use of VR in the terms of the learning it purports to facilitate.

Whilst the method required the use of expert participants drawn from teaching staff who were familiar with the simulators, it considered the time commitment and took steps to keep this to a minimum. It used an efficient think-aloud protocol coupled with a video recording to capture as much information as possible in a single session. Formalisation of the individual descriptions was carried out without participant input, only seeking optional comment on the written-up individual task description. A group-consensus interview carefully walked all participants through the draft consensus task description compiled from the individual descriptions and provided a further opportunity to comment.

The resultant task description from any analysis must be suitable for its intended purpose. Here, the description will inform the development of exercises for VR simulation that focus on transferable cognitive aspects of the task. In this regard, the analysis has captured an adequate description of the main areas of activity, the attentional focus, common mistakes and evaluation criteria, plus, goal-specific limitations of the current simulation environment. The task description is not the end point of developing educational resources and this approach has provided a rich source of material from which exercises and learning materials can be developed. Therefore this approach is considered suitable for its purpose.

To guide the identification of the most appealing training opportunities and the development of these exercises, the next chapter will discuss the task description produced by this approach in the context of the psychomotor skills literature with reference to the limitations identified by the participants.

Chapter 6

Exploring the Task Description

This chapter will explore the consensus task description (Summarised in Figure 5.5) through the lens of the motor skills literature with the goal of identifying opportunities for training in the VR environment. Practice in the VR context is not done for its own sake, it is done with the intention that the skills and knowledge developed will transfer to a criterion task, namely the demonstration of the skill in another training environment or in real clinical practice. This exploration is a continuation of the instructional design aspect of Task Analysis because it is the responsibility of the analyst to identify the most promising training opportunities from the task description and present them in a form from which a novice can learn (Annett et al., 1971).

Furthermore, this chapter begins to draw the distinction between task decomposition and deconstructed exercises. The process of decomposition, through Task Analysis methods, has revealed and documented the steps and 'abstract' underlying knowledge associated with the selected operative procedure. To produce learning material from this representation these concepts must be reified through the identification, design and creation of cognitive-focussed *deconstructed* exercises. The content of these exercises can be combined and re-constructed by the learner to form an understanding of the whole task. This terminology distinguishes the process of breaking down and describing a real-world task from the conscious design of novel component-part exercises that teach the cognitive knowledge revealed by the analysis.

The full Task Description is presented in Appendix D, here, each goal will be related to the VR simulation modality, discussing issues identified by the participants, any problems that these create, and promising areas for future exploration as deconstructed exercises.

6.1 Exploration of the described goals

6.1.1 Preparation

The Preparation goal consists of the steps required to safely commence the operative intervention. This goal, and its sub-goals, presented difficulties for the participants to describe what would be done at this stage because significant differences exist between the VR setting and the real world. These differences resulted in participants having different interpretations of the pre-task narrative and what was relevant to describe. This is a known risk of the use of simulation (Kirwan and Ainsworth, 1992, p. 151), but for this study it led to a broader consideration of the task and revealed aspects that are not present in the simulation or are modelled in such a way as to alter how the goal is met. As discussed below, these differences may lead to undesirable negative transfer when performed in other contexts.

Assessment

Beyond the motor activity of the operator placing themselves in the correct position to achieve visibility of the tooth requiring treatment, the assessment subgoal is predominantly described as a cognitive activity supplemented with affective domain skills. The identification and confirmation of the presence of caries visually and from the radiographs is predominantly based on visual cues drawing upon knowledge of the characteristics of the appearance of demineralised tissue. Pre-operative assessment is an important skill for the learner to master because it is the goal that determines if operative resection is indicated or if other less intrusive interventions are appropriate.

The VR simulation environment used for this study does not currently present assessment as an explicit step, nor are simulated radiographs routinely presented to the learner. A tool that provides cross-sectional views is available but this does not have a practical real-world equivalent. This tool permits 'flying' through the tooth to view all aspects of where the caries is present as 2-dimensional slices, whereas a pre-operative radiograph presents a projection of 3-dimensional object on to a 2D plane of the radiopacity of the exposed tissues. The enhanced visualisation available from the cross-sectional tool could be viewed as introducing construct-irrelevant easiness (Messick, 1995) as it makes this aspect of the task simpler than the corresponding real-world activity and also misses an opportunity to develop skill in radiographic interpretation.

Reviewing radiographs is a necessary step of assessment and pre-operative planning, therefore, its exclusion from the VR task could result in undesirable consequences. For example, if radiographic examination is not afforded significance in the simulation environment, by what means does it acquire the correct significance in the operative context?

How does 'skipping' this step here prepare the learner for the real-world context where this represents a step in the expected performance?

Without guidance from a tutor, the present system does not prevent or discourage the learner from loading an activity and commencing the task without any consideration of relevant factors: that treatment is indicated; that the proximity of the caries does not pose a significant risk to exposure of the pulp, nor any patient factors that should be influencing their decision making process. Whilst an exercise can usefully simplify the problem to reduce the cognitive load in order to aid learning, here this is a simple omission rather than a conscious attempt to enhance the learning experience. Omitting crucial steps in simulation environments has been shown to have disastrous effects in aviation (Weick, 1990) so the removal of what should be positive actions in the real situation should be done with caution lest it diminish their perceived value and result in undesirable habit-forming.

It could be possible to incorporate an assessment step into the simulation experience. This could permit consideration of the above pre-operative factors that lead to the affirmative decision that treatment is indicated. As a predominantly cognitive focussed step it is likely that skills developed here would favourably transfer to other contexts (Anderson et al., 1996) and as such would be worthy of exploration. However, in the context of this work this goal precedes the operative intervention so adding functionality to the simulator itself to support a pre-task activity may not be the optimal allocation of effort when it could be achieved by a tutor-led presentation/discussion prior to starting the task.

Set-up simulator

The goal of setting up the simulator involves rotating the virtual camera to a position at approximately '10-to' on a clock face relative to the exercise and zooming in to a magnification of approximately equivalent to 2x magnification loupes. If needed, the height of the simulator and operator's chair can be adjusted at this stage and if necessary the operator can calibrate the virtual handpiece to align its relative real-world position in order to achieve a safe finger-rest on the finger-rest stage.

There is no directly equivalent step in the real world procedure, although a weak (albeit out of sequence) parallel can be drawn to the setting of the height of the patient's chair and the operator positioning themselves to commence the procedure. Given the nature of the virtual reality modality, it is inevitable that some concessions to the limitations of the hardware are inevitable in much the same way that inserting the prepared arch into a phantom head has no real-world treatment parallel. Given this study's goal, the documentation of these limitations is of interest and worth documenting.

To achieve this goal, the operator must be aware that the environment is not optimally configured via their physical relationship and proprioception. However, adjusting the simulator's height and ensuring a secure finger rest is driven by an underlying awareness that these parameters should be addressed prior to commencement of the procedure. Unfortunately, current simulators are unable to detect these external positional factors (Towers et al., 2019) so cannot provide corrective feedback to ensure they are attended to. Furthermore, many of these corrections would be specific to the VR modality and would present differently in the real world. Therefore, whilst the learner should be mindful of these considerations, practising and perfecting their performance would not usefully transfer to the real-world task. It could be a useful addition for the simulator to remind the operator of their musculoskeletal health and the importance of a secure finger rest, however, it is not a skill that can be readily developed using simulator exercises.

Select appropriate burs

This goal differs significantly from what would occur in the real-world equivalent task. Within the VR environment, a user is able to simply select a handpiece and choose burs by tapping their corresponding icon on the simulator's touchscreen. In the clinical context, the operator must pro-actively retrieve a selection of appropriate burs from a dispensary or storage location so that they are conveniently to hand during the procedure. This step requires the operator to preempt the appropriate design and sizes of burs that will be necessary to cut a cavity of the desired size and shape and then retrieve these (or the nearest alternative) from the dispensary. No consideration of this activity was present in the VR context used for this study and the user is presented with a comprehensive list of all available burs.

Participants described that learners should be aware of the real-world costs of switching between burs. Many burs are single-use and others require thorough cleaning between uses, these both carry a financial cost so the indiscriminate use of a large number of different burs of different designs is viewed as an undesirable behaviour. Furthermore, changing burs takes time and there is an increased risk of cross-infection risk with every change so the consequence-free instantaneous switching between an exhaustive selection of burs may give unrealistic expectations of the real-world procedure or develop undesirable behaviours that must be unlearned.

Within reason, the choice of burs for any given task is a mater of personal preference. Different operators may prefer slightly different design burs for the same task drawing on their own training and experience (Mansueto et al., 2010). The cognitive knowledge that a bur is inappropriate can be trained, however the individual decision relating to preference within those boundaries is subjective and therefore not an ideal task for VR training. This is compounded by the variance in the behaviour of the simulated tools and their real-world equivalents (Towers et al., 2019) meaning that any preference developed may be built upon behaviour that is not representative of real-world performance.

Therefore this task may be more suited to training elsewhere.

6.1.2 Access to Lesion

The overarching objective of the Access to Lesion goal was described as to gain minimal access to the cavity such that the infected dentine is accessible and the ADJ is visible.

Access cavity through enamel and Establish initial cavity outline

These two related sub-goals both require the demonstration of fine finger-dexterity to accurately introduce the bur to the specified depth and create the initial cavity outline. As expertise develops, the two goals become a single operation which is consistent with motor skills literature that suggests that performance becomes smoother and the inter-task boundaries begin to blur as performance becomes a procedural ability (Tjiam et al., 2012). But, for a novice, a focus on building a meaningful mental model of this goal and explicitly directing their attention towards the factors they should be attending to at a sub-goal level may develop a more transferable representation of the skill.

The motor skill aspects of these sub-goals suggest that repeated practice may be of limited value, however, it is underpinned by cognitive knowledge and there are opportunities to direct attentional focus and emphasise cognitive aspects of the task in order to assist avoiding the most common mistakes in its execution. These include:

- Recognising appropriate places to access the cavity
- Recognising the differences between the presentation of carious and healthy tissue
- Recognising that the ADJ has been successfully accessed
- Where to direct attention so that the correct depth and size access cavity is created
- Demonstrating the patterns of the brushing motion to be used when extending the access to form the initial outline
- Identification of the long axis of the tooth and approaches to maintain the correct angulation

Any educational activities that contribute to the awareness of these underlying attributes of performance could be a valuable addition.

Inspect

The Inspect subgoal provides an interim operative inspection and requires the application of cognitive knowledge to holistically evaluate previous performance to establish if it is appropriate to progress to the next stage or if earlier goals should be re-visited.

The activities performed in this step differ somewhat from what is done in the real world and are impacted by limitations in the fidelity of the simulation experience, including:

- The affordances of the VR modality encourage the user to rotate the view to inspect the periphery of the cavity. In the real world, the mirror would be used to inspect some internal walls of the tooth so this disagrees with the expected behaviour which risks diminishing its value or developing a habit of non-performance in the real context.
- The tissue colours used in the VR simulator for this study give an impression that the differences between tissues is more visually pronounced than is the case in the real world. The lack of nuance may lead a novice to only expect, and consequently look for, stark differences in presentation when the real-world distinction is more subtle.
- There is no ability to wash or dry the preparation in the VR context. Drying the preparation reveals information about the carious spread. If this is not attended to in the training environment it is unlikely that this skill will be developed to be deployed in other contexts and may form a habit of not doing so.
- Finally, the VR context permits drilling from any physical angle (within the operating envelope of the haptic arm). This means that the tooth may be prepared in a way that is not physically possible in the real world and result in the learner expending thought and reflections on a performance that are meaningless outside of the VR context.

With the limitations presented above, it is unlikely that a learner trained exclusively using VR would attend to all aspects in performance of the real task. If an attribute of performance is not available in the training context, the learner can't be expected to attend to that same attribute in the real world so it is much less likely to transfer (Schmidt and Lee, 2014, p. 189). It could be possible for a tutor to proactively inform the learner that certain aspects are not fully represented in the simulation, however, it is uncertain if this 'appended' stimulus would be integrated in the same way and may simply manifest as a source of extraneous cognitive load for the learner.

This goal is predominantly a display of cognitive domain activities, so exercises that draw attention to the factors that should be inspected may hold promise for effective transfer. However, the simulator fidelity limitations may limit what is achievable using the VR simulators available at the host institution for this work.

6.1.3 Establish Caries-Free Margin at ADJ

Participants noted the importance of establishing a caries free margin at the ADJ. Its successful execution relies upon cognitive factors of perception and it is an iterative and highly process-driven goal suggesting that it holds promise for instruction using deconstructed activities. The present whole-task simulation has elements of construct underrepresentation (Messick, 1995) as participants reported that the simulation fails to restrict viewing angles, simulate water cooling, the use of an aspirator, rubber dam, varying levels of demineralisation or any factors relating to any soft tissue management. Their absence here contributes to an unrealistic impression of the visibility available when carrying out this task in the real world context. If these issues cannot be readily addressed, then alternative exercises that guide the learner in establishing the declarative base knowledge of the steps and success criteria involved could result in superior knowledge acquisition that could be demonstrated elsewhere.

Select appropriate bur

As discussed above, the comprehensive range of simulated burs presents issues for their proportionate use. The criteria that inform the selection of an appropriate bur for addressing caries at the ADJ are items of declarative knowledge, and a poor selection can result in undesirable consequences. Therefore, these aspects and their consequences could be introduced with specifically focussed exercises. However, developing this awareness may be better achieved outside of a VR context where the bur's behavior is more representative of its real world performance.

Remove accessible caries at ADJ

The execution of the tasks in this goal draw upon a great deal of cognitive knowledge: an awareness of appropriate instruments, their angulation, the nature and appearance of the tissue to be removed and its behaviour whilst doing so. All of these aspects contribute to the decision making that drives motor performance. Common mistakes involve selecting inappropriate instruments, their inappropriate application and prematurely advancing to later steps. Avoiding these mistakes requires the application of cognitive knowledge to guide appropriate motor performance and decision making in a dynamic environment that changes in response to previous actions.

This analysis reveals a limitation in the current simulation because colour and texture of the tissues being removed are not fully modelled. There is no representation of the soft and readily removed tissue transitioning through a 'leathery texture' to a firmer and dusty appearance. Nor does the simulated probe provide the correct sensations when checking for the presence of non-visual cues of demineralisation. These factors underrepresent the scenario and preclude directing the learner's attention to aspects of performance that guide and inform expert decision making in the real-world task. Again, in the absence of these cues the learner cannot be expected to attend to them in other environments as a result of their VR simulation training.

Participants reported that they were compelled to incorporate proxies and estimates from experience to guide their judgement and inform their actions. However, for a novice lacking this experience, the decision making process can only be informed by what is presented in the simulation and any absences mean that an incomplete mental model of the task will be formed that is not fully reflective of the real-world presentation.

The fidelity of the simulation could be enhanced over time to address these shortcomings and present these attributes, however, their simple presence does not draw the learners attention towards them or recognise their significance to the decision making process. Again, drawing attention to an attribute's significance is left as a task for the tutor. However, a series of exercises that point out the features that should be attended to and develop the methodical application of the correct process may result in a more transferable skill than simply representing a more realistic carious tooth.

Inspect periphery

Participants noted that novices often fail to identify that caries remains at the ADJ due to incomplete checking or failure to recognise that unsupported enamel was impairing their view. As described, this goal does not rely upon motor skills but it does require that attention is focussed in a methodical process to ensure that it is completed successfully. As discussed above, limitations in the current VR exercise's representation of the carious tissue means that not all factors that should be attended to are represented, resulting in potential limitations for the development of this skill. However, participants noted that some learners fail to carry out this step *at all* so perhaps exercises that illustrate that an inspection is required (and emphasise its importance) may result in habit forming and transfer.

Manage unsupported enamel & regain access to ADJ

The success of this goal hinges on careful and fine motor control to remove the unsupported enamel and regain visibility of the ADJ. These motions are directed by an awareness that the step must be carried out and a recognition of the indications that access to the ADJ has been restored. Clearly explaining and demonstrating the gestures required to remove the enamel and presenting exercises that clearly explain what it means for the ADJ to be inaccessible could guide performance in such a way that promotes transfer to other contexts.

Evaluate ADJ

The final sub-goal is a further inspection and an evaluation of the ADJ. In this goal, the newly accessible and visible areas of the ADJ revealed by the management of the unsupported enamel are inspected to see if further caries must be removed. The analysis documented that a great deal of the checking and judgement required for this goal is carried out using hand tools. The VR simulator used for this study lacks physical or software-simulated representations of most hand tools used in caries removal. Furthermore, the probe, whilst present, does not faithfully reproduce the tactile sensation of probing carious tissue. These limitations mean that the operator does not and *can not* use the same process in the VR context as would be used when carrying out the real world procedure. Therefore, there is no opportunity to emphasise the correct behavior or to use the appropriate instrumentation in the VR environment. The absence of these instruments may erroneously diminish their importance or reduce the learners recognition of their relevance to this stage.

The VR simulator used for this study simplifies the representation of demineralised tissue to a single solid brown colour which creates context-irrelevant easiness in the execution of the task. Effectively, the only aspect a learner is required to attend to is the removal of all of the visible brown voxels. The real world procedure is more nuanced, requiring judgement and cognitive knowledge. Real world caries visually transitions in line with the level of demineralisation and in chronic caries can have a long transition back to healthy tissue creating uncertainty as to when to stop extending. The simplification in the VR environment used removes the need to be aware of this as successful performance in this context simply requires that the brown material is removed.

This construct underrepresentation means that the learner is not prompted to direct their attention to a change in the appearance and texture of the tissue, nor are they prompted (or able) to select appropriate equipment to judge these factors. The lack of accurate haptic feedback and the incomplete list of equipment may mean that exercises cannot be developed to fully train this aspect of the procedure. The differences and limitations may prove to be too great to result in reliable transfer in this case.

6.1.4 Manage Caries at Pulpal Floor

The successful performance of this goal depends upon successfully selecting the most appropriate instruments, using them correctly and giving consideration to case specific factors. As discussed above, the current realisation of the simulation software represents does not fully model all attributes of carious tissue at the varying degrees of demineralisation. This prevents the learner from applying the relevant cues to inform and evaluate their progress. Due to these limitations, participants noted that their current teaching practice in the VR context was to encourage learners to simply leave some caries in the base of the cavity as a token representation of affected tissue. This performative act is used to demonstrate an awareness of the infected-affected distinction and a higher degree of handpiece control. However, the basis of this action is not rooted in the application of best practice - the learner is simply retaining caries in the place that they were asked to. Additionally, hand tools are frequently used at this stage so their absence represents a further construct underrepresentation and results in a step that cannot be carried out in VR in the same way as in the real world.

Similar to the *Remove accessible caries at ADJ* goal above, its execution combines underlying cognitive knowledge and motor skills. Exercises to learn and apply the principles and processes could be developed using a deconstructed approach, however, due to the lack of hand instruments these lessons would be incomplete and only represent a subset of the actions carried out in the real world. This makes this goal less appealing for further exploration.

6.1.5 Refine Cavity & Prepare for Restoration

This goal, whilst distinct in its objectives, overlaps significantly with the previous goal in its execution. It is comprised of the subgoals of refining the internal form, managing the margin and a final review prior to restoration. However, a skilled operator will be giving attentional focus to planning ahead and considering factors relating to the final design of the cavity whilst carrying out the previous steps, making small adjustments to the way the caries is removed such that the final design *evolves* towards the desired form rather than these features being added with separate and discrete actions. However, a novice can benefit from specifically considering this goal to ensure that the prepared cavity meets the design requirements.

Refine internal cavity form

This goal requires that the learner applies their cognitive knowledge of the principles of cavity design to what they have produced. If any errors are detected then these should be corrected and the internal cavity shape refined. The principles of cavity form could be taught in a novel way using deconstructed exercises developed for the VR context. The most common mistake reported at this stage was over-refinement of the cavity so there may be value in providing focussed instruction on how this step is best approached to minimise these errors.

Remove unsupported enamel, refine cavo-surface angle & margin

Whilst checks for these attributes will have been performed throughout earlier goals, participants described the importance of carefully ensuring that no further adjustments are required. The most common mistake by learners at this stage is to fail to carry out this step at all. This could arise from a lack of awareness of its importance so this could be an opportunity to direct attentional focus towards it and its role in the production of a high quality cavity preparation. There may be potential for developing exercises that illustrate its place in the task and the consequences of failing to carry it out which may promote a transferrable awareness to other contexts.

Final review & ensure suitability for restoration

At this point the cavity preparation can be considered complete, however, participants described that one final review inspecting all aspects of the preparation should be carried out as a conscious and deliberate step to ensure all design criteria have been met. These checks represent an application of all relevant cognitive knowledge to ensure that the previous goals have produced a satisfactory result. Any shortcomings would require the operator return to the earlier step and follow the guidance to make any necessary adjustments or corrections.

As this is predominantly a final verification of the same checks as performed in previous goals, any exercises developed to instruct the learner are likely a repetition these. Therefore, instruction for this goal could be represented as a combination of these existing exercises.

6.1.6 Critique

The final goal identified by participants was to critique their outcomes. The previous goal ensures that the restoration is fit for purpose, but here the operator is asked to return to the pre-operative radiograph and identify aspects of the performance that went well or badly and to reflect on any contributing factors. Attending to performance and deliberate reflection is a good educational practice (Lin et al., 1999) and is a graduate attribute required of dental students (GDC, 2015).

Participants noted that the VR simulation environment has the potential to reveal additional information beyond what is normally available such as high magnification and cross-sectional views to permit opportunities for further critique. However, this must be balanced with an awareness that the fidelity offered by these views can exceed the level of clinical significance. This factor could combine with the ability to easily reset the exercise for another attempt and change the nature of a learner's reflection to one where an irrelevant degree of perfection is sought and mistakes are quickly attributed to motor performance. This contrasts starkly to the real world experience where adaptability is often required to achieve an acceptable result in the face of the unchangeable consequences of external factors or poor decisions.

As the goal of this step is internal reflection, interactive VR simulation exercises are perhaps not the optimal environment to achieve this.

6.2 Considerations for VR dental education from this Task Analysis

Using VR to conduct a task analysis of an operative procedure has produced interesting results. Participants describing the task were forced to accommodate the limitations of the simulation itself and expend effort upon explaining how it differs from what they would do in the real world. Decomposing tasks via a task analysis method identifies skills that are not observable in expert performance but are precursors to acquiring the skill being demonstrated (Frederiksen and White, 1989). Attentional focus is a key attribute of expert performance (Schmidt and Lee, 2014, p. 64) and this analysis has demonstrated that the simulators used here lack numerous key visual, tactile and instrumental considerations that should be present for the development of the skills necessary for the criterion task. This supports the idea that alternative uses should be considered that may be more effective at promoting transfer of skills acquired in the VR context.

The absence of instruments, differences in the way they behave, differences in the tactile and visual presentation of healthy and diseased tissue and the removal of patient factors all contribute to significant changes in the way procedures are performed and the ability to reason about the factors that inform decision making. Some differences may be unavoidable to accommodate technical limitations and necessitate simplifications that can scaffold future learning. However, this task analysis has revealed that the VR experience differs so greatly from the real-world experience that, in some cases, completely different actions must be taken. An experienced user may be able to acknowledge these limitations and gain other benefits from the simulation, however, "what you learn depends largely on what you practise" (Schmidt and Lee, 2014, p. 200) and a novice may not fully appreciate that they are experiencing something that differs from the real scenario. This could lead them to attend to factors that would never manifest in the real world or neglect factors that should receive their attention resulting in extraneous load when they later try to reconcile and transfer their learning to other contexts.

This analysis further demonstrates the limitations of the feedback mechanism currently offered by many simulators. As discussed, many factors of performance are continuously being applied and the performance judged against expectations and measurements taken from visual and tactile sensations. In current scoring mechanisms, this rich multi-sensory guidance that informs the decision making process is reduced to a simple score derived from from target-region removal. This feedback measure is not representative of the process used operatively nor does it correlate with any of the cues that the learner should be attending to in order to evaluate their performance and promote improvement (Dixon et al., 2021).

More broadly, and through the lens of the validity literature, it could be argued that the omissions and limitations in the scoring mechanism prevent any fair measure of a student's performance. Expert performers discussed how absent attributes modified their approach so a novice trying to reconcile the VR experience with content taught elsewhere may be misled and make erroneous decisions as a result. A low score should never result from an assessment that does not allow a learner to demonstrate their true ability because a particular aspect of performance was underrepresented in the task (Messick, 1995). The validity of a test cannot be detached from its intended purpose (Messick, 1989, 1995) but based on this task analysis, the simulation may not be considered a sound basis for making inferences of performance (Cronbach and Meehl, 1955) so their use for any credit or consequence bearing assessment should be avoided.

These same limitations impact the measurement of skills transferred from the simulation context. Transfer is often investigated using a 'first-shot' measure. This compares how the learner performs during their 1st attempt at dealing with a situation after undergoing training (Hammerton, 1967) and any difference in performance can be attributed to that training. However, the limitations discussed above mean that the learner has not actually been exposed to many of these relevant stimuli during training so is effectively encountering them for the first time, rendering any first-shot measure of the criterion task unreliable.

It is not unreasonable to conclude that the whole-task simulation of operative procedures emphasises these deficiencies and confounds the learner's ability to construct a complete and accurate mental model of the task. Specificity of learning is the dominant characteristic in the training of motor skills (Schmidt and Lee, 2014, p. 200) so practising a whole dental-operative task on a VR simulator with the hope that it will improve motor performance at the real-world task is likely to produce disappointing results. The next section will present deconstructed part-task training exercises derived from the task description that focus on the cognitive aspects of operative tasks rather than their whole-task simulation. The use of these exercises can then be compared with traditional whole-task training to explore if they result in superior transfer to other contexts and suggest a more effective paradigm for the use of VR in pre-clinical training.

Chapter 7

Developing Part-Task Deconstructed VR Simulation Exercises

Learning a complicated motor skill can be challenging. The more complex the task and the more sub-goals it contains, the more likely it is that it will exceed the processing capacity of a novice to perform it (Paas and Merrienboer, 1994). Part-task activities are one way in which this complexity can be managed and it is recommended that they are integrated throughout the learning process to help students to acquire skills (Frederiksen and White, 1989; Schmidt and Lee, 2014, p. 219).

The use of part-task practice reduces the likelihood of novices wasting energies on understanding task-irrelevant details and focusses their attention to the aspects that facilitate schema acquisition. This should reduce the time and mental effort required to acquire a skill than would be required when using conventional approaches (Paas and Merrienboer, 1994). The practice types which have been shown to be most successful for the training of complex cognitive skills include the use of worked examples and a study assignment (Paas and Merrienboer, 1994). Therefore, this is the broad approach that will be adopted below.

Whilst breaking down a task for part-practice has shown to be ineffective for certain motor-skill focussed activities (Schmidt and Lee, 2014, p. 220), manipulating the process by which schemas are acquired for complex cognitive tasks should be encouraged to aid their acquisition and transfer (Paas and Merriënboer, 1994). Worked examples follow this guidance and are regarded as more natural learning material than explanatory texts so their greater use often leads to improved learning outcomes and transfer (Sweller et al., 1998). The interactivity that VR systems afford means that they have the potential to deliver the cognitive knowledge that would previously be provided via text books, but do so in a way that will promote enhanced schema acquisition via interactive worked examples.

This chapter will describe how the task description detailed in the previous chapters was used to create a series of activities that facilitate exploration of the cognitive underpinnings of caries removal and cavity preparation.

7.1 Cognitive Aspects of the Task Description

The task description produced in previous chapters provides a detailed account of the selected operative procedure. It describes the procedural elements of the task, where attention should be focussed and the motions to required to carry out each step. It also details the attributes of successful performance, common mistakes made by novices and itemises where the simulator's representation of the task diverges from that encountered in the real-world. To assist the development of the exercises and activities, the cognitive aspects of the task were extracted in the form of declarative statements which are presented in table 7.1. Many aspects of cognitive knowledge are reused throughout the execution of the task, however, for brevity the table lists each aspect only against the first goal in which it is required.

Table 7.1: Underlying attributes of cognitive knowledge

Goal	Cognitive knowledge
Preparation	Knowledge of : • The appearance of caries during visual inspection • Ways in which a patient's treatment history can influence the approach to treatment and any necessary adaptations • How to interpret a radiograph in order to estimate and visualise the size and extent of a cavity • The indicators from a radiograph that suggest operative intervention is necessary • The implications and risks associated with caries which has progressed to be near to the pulp when seen on a radiograph • The correct way to hold the handpiece and other instrumentation • How to orientate oneself and position the patient to ensure the optimal viewing angles and an ergonomic posture • The importance of a secure finger rest and how it can be established • The purpose and function of different burs so that appropriate selections can be retrieved prior to commencing the procedure • The appropriate handpiece and bur to use when accessing the lesion

Continued on next page

Table 7.1: *(Continued)* Underlying attributes of cognitive knowledge

Goal	**Cognitive knowledge**
Access to Lesion	Knowledge of : • How to integrate information from the radiograph and the visual presentation of the lesion to determine the optimal access point • The correct bur angulation when accessing the cavity • The morphology of the tooth in order to anticipate the tissues that will be encountered at any given depth and assist with avoiding iatrogenic damage • How to use landmarks on the bur and tooth to estimate and maintain consistent depth (and the corresponding knowledge of what an appropriate depth is for this stage) • The nature of the motions to make with the handpiece and the cutting properties of each bur in order to remove carious enamel • How to identify caries that has progressed through the enamel to the dentine and the appropriate corresponding extent of the initial cavity outline • The tendency for depth control to waver and the corresponding importance of active monitoring of the depth and angulation in order to maintain an even depth • The appearance and approximate location of the ADJ and how to recognise when it has been accessed

Continued on next page

Table 7.1: *(Continued)* Underlying attributes of cognitive knowledge

Goal	Cognitive knowledge
Establish Caries-Free Margin at ADJ	Knowledge of : • The visual appearance of caries at the ADJ and the associated non-visual tactile sensations when determining the extent of demineralisation • The presentation of sound tissue in order to determine that the limit of the caries has been reached and the ADJ has been successfully cleared • Criteria for selecting an appropriate size of bur for the removal of caries at the ADJ • The appropriate depth of the caries free margin to be cleared at the ADJ • The need to manage pulpal floor caries separately and to address it separately during a separate goal • The visual appearance and texture of the tissue being removed and how this relates to the level of demineralisation • The correct bur angulation and motions used when removing caries at the ADJ with the low speed handpiece • The use of hand tools to assess the density of tissue and how these behave when interacting with tissue of different levels of demineralisation • The methodical one-internal-wall-at-a-time approach to manage the complexity of the task • The appropriate handpiece to use for each task within this goal • How to recognise that enamel has become involved and is impairing visibility of the ADJ • How manage the removal of unsupported enamel and recognise when this has been achieved

Continued on next page

Table 7.1: *(Continued)* Underlying attributes of cognitive knowledge

Goal	Cognitive knowledge
Manage Caries at Pulpal Floor	Knowledge of • How to integrate information acquired during the preceding steps with the radiograph to re-assess the risk of pulpal exposure • Partial and total caries removal approaches and the indicators for choosing between them • The appropriate bur and handpiece to use and the correct motions to use whilst removing caries from the pulpal floor • How to discriminate between caries at different levels of demineralisation • The criteria for the retention of affected tissue during partial caries removal • How hand tools are used when removing caries and judging levels of demineralisation • The expected form and attributes of a well prepared cavity base

Continued on next page

Table 7.1: *(Continued)* Underlying attributes of cognitive knowledge

Goal	**Cognitive knowledge**
Refine Cavity & Prepare for Restoration	Knowledge of • The minimum width of enamel that must be present around all walls of the cavity so that if this design goal cannot be met an alternative more appropriate design can be proposed • The criteria necessary for the cavity design to ensure compatibility with the intended restoration material and how to accurately judge that this has been achieved • The desirable attributes of the cavity floor and walls • Appropriate management of unsupported enamel and desirable attributes of a well prepared margin • The management of caries in order to judge that an appropriate volume of carious tissue has been removed and no caries has been inappropriately retained • Clinical acceptability criteria to evaluate the preparation's suitability for restoration
Critique	Knowledge of • The criteria of the *ideal* preparation • The clinical relevance and implications of any competing priorities of the ideal preparation when applied to this specific case • What was achievable given the patient's history and specific case presentation and how to reconcile any differences between what was produced and the ideal to inform future practice

Table 7.1 shows a substantial range of declarative knowledge items that underpin this skill. If only a single activity to address each item were developed it would still result in a quantity of activities beyond what is feasible to produce and test within the scope of this thesis. Therefore, a reasonable subset of goals must be selected to explore the deconstructed approach. If activities produced to teach this subset result

in positive results and transfer then additional exercises can be added in the future. A stage that participants identified that learners find the most challenging is a good candidate for selection as it will lead to the greatest impact. Unfortunately, the task analysis prioritised identification of common mistakes rather than a subjective ranking of the relative difficulty across different aspects of performance. However, many such observations were shared incidentally during the individual task analysis sessions so these were available after re-reviewing the recordings. Explicitly prompting for this question for this would be a good addition to the standard task deconstruction questions for any future task analysis.

The aspects of the task identified as where students struggle the most were when determining how (and where) to *Access to Lesion* (Goal 2) and the removal of appropriate tissue in order to *Establish Caries-Free Margin at ADJ* (Goal 3). These two goals are sequential in the task description, so a series of exercises could be developed covering both, however, much of Goal 2 is derived from information acquired from a visual examination and relevant radiographs. Limitations in the fidelity of the simulated representation coupled with much of the information being outside the simulated environment leads to limited opportunities to develop interactive exercises without significant re-structuring of the simulation environment itself. Therefore, the focus of the exercises developed here will be to explore concepts relevant to establishing a caries free margin at the ADJ.

Much of the process for creating these exercises is a creative endeavour (Jonassen et al., 1999, p. 4). The information extracted from the task analysis in Table 7.1 presents a series of key instructional outcomes that should (where supported by the simulator) be incorporated. However, the specifics of how this done and where or how each component is addressed plus any resultant effectiveness is somewhat dependent on the skill and creativity of the analyst (Drury, 1983). This does present an issue for the reproducibility of this approach; it is conceivable that different analysts would derive different exercises from the same source material. However, the comprehensive task analysis process used should ensure that all educational outcomes are addressed (Jonassen et al., 1999, p. vii) and that any differences are comparable to any other inter-tutor difference in teaching approach.

7.2 Exercises to support the instruction of underlying concepts of the "Establish Caries-Free Margin at ADJ" goal

The *Establish Caries-Free Margin at ADJ* goal is a mostly self-contained activity with few external dependencies for which a cohesive suite of educational activities can be created. The remainder of this chapter will describe the development of a suite of exercises related to the underlying concepts of the above goal, as listed in row 3 of

Table 7.1. To resolve a dependency of this goal, additional activities will address an awareness of the appearance and location of the ADJ which would otherwise be covered by activities for the previous goal.

The tasks of this Goal can be grouped into 3 conceptual areas:

- An appreciation of the location of and access to the ADJ
- The presence of decay within a tooth; how the decay in different locations is referred to and how tissue affected by the carious process transitions from infected to affected levels of demineralisation.
- The use of a methodical approach to clearing caries at the ADJ using appropriate instruments at each step.

When using a dental VR simulator in a classroom setting or to facilitate an educational intervention there is a time cost associated with selecting exercises and waiting for them to load. This has implications for the practicalities of conducting an experimental intervention as including more exercises will require a greater time commitment from volunteers which may reduce the likelihood of participation. Therefore, the design of the exercises below will (as much as possible) minimise this switching by including multiple activities within a single exercise, using the above conceptual areas as an initial grouping.

However, the number of activities that can be represented within a single simulator exercise is not limitless. The computing resources available to the simulators are finite: complex scenes place a greater demand on system resources and any additional activities within the same scene will require proportionately more resource. If an exercise is of such complexity and size that it exceeds the available resources, tradeoffs must be made between resolution and frame-rate (Claypool and Claypool, 2009). The simulators available for this study have the resolution locked by the vendor so the impact of this trade-off will be most acutely felt in the frame-rate. Studies investigating the impact of frame-rate on user satisfaction for video content and have found a high tolerance for frame-rates as low as 3 frames per second, however the interactivity found in video-games (and by extension computer based simulations) means that a substantially higher frame-rate is critical for adequate performance (Claypool et al., 2006). When the reduction in frame-rate introduces lag in the representation of hand tracking (as would be highly noticeable in a dental simulation) this is associated with a considerable performance degradation (Ware and Balakrishnan, 1994) and is likely to adversely impact upon the utility of the exercises. Therefore, a balance must be struck between the creation of exercises that are internally consistent and fully explore any given concept and splitting them into separate exercises to manage their computational complexity to mitigate any performance detriment.

Large and high-resolution models containing many tens of thousands or millions of vertices require proportionately more system resources than simpler lower resolution models. As discussed in Chapter 2 the need for a truly realistic visual aesthetic has not been demonstrated therefore, simpler lower resolution models will be utilised for this work. These will reduce the demand on the simulator hardware, releasing system resources to include more activities in a single exercise.

The following subsections will describe 4 exercises, each containing one or more activities that explore the underlying knowledge necessary to complete the goal of establishing a caries free margin at the ADJ in an occlusal lesion. These new exercises will use simulator functionality developed specifically to support this work[1]. Each exercise will begin with instruction to orientate the learner with the task. As the learner interacts with and removes material context-relevant prompts will pop-up and direct their attention and provide further instruction on how to continue. Knowing where and when to focus attention is a key discriminator of expert performance (Schmidt and Lee, 2014, p. 64), so by delivering context-aware instruction during exercises this might enhance the acquisition of this knowledge. These notifications will appear when a threshold percentage of specified tissue has been removed. The threshold values that trigger the display will be determined empirically during testing of the exercises running on the simulator.

Full details of the implementation of each exercise are provided in Appendix E

7.2.1 Exercise 1: Developing an appreciation of the location of and access to the ADJ

Clearing the ADJ is the focus of this goal, however, without an understanding of what the ADJ is or where it is located this will be hard to achieve. Therefore, the first exercise will cover the following related concepts from Table 7.1:

- The appearance and approximate location of the ADJ and how to recognise when it has been accessed
- The correct bur angulation and motions used when removing caries at the ADJ with the low speed handpiece
- How to recognise that enamel has become involved and is impairing visibility of the ADJ

[1]To support the goals of this work, HRV (Laval, France) added additional functionality to the simulator to enable attentional focus to be directed. This included: 1) The ability to set arbitrary colours on imported .stl files (previously only 4 colours corresponding with dentine, enamel, pulp and caries were available) 2) The ability to display pop-up notifications to the user after a threshold percentage of an imported .stl had been removed

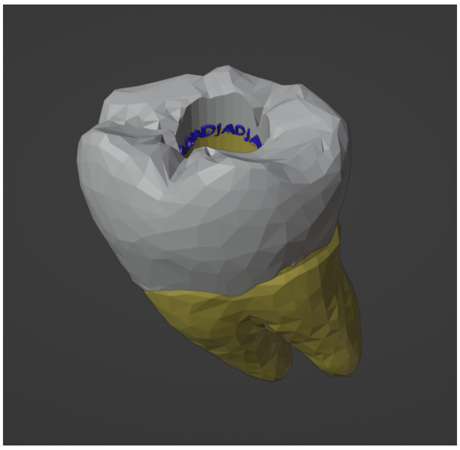


Figure 7.1: Exercise 1 Activity 1, Location of the occlusal ADJ

To cover these concepts, this exercise will present the learner with 3 activities. Introductory information will describe basic concepts of bur angulation, seating to the appropriate depth and how to utilise the bur's cutting sides and then instruct the learner to interact with the presented activities from left to right. A future implementation could move these introductory concepts into a separate, more exploratory, exercise but for the purposes of this work careful participant selection can ensure an awareness of these concepts and the allow the information to simply serve as a reminder.

Activity 1.1

The first activity introduces the location of the ADJ in the occlusal aspect. A tooth will be displayed with a stylised access provided. Learners will be asked to widen the access cavity without going deeper into the tooth. Lettering will be shown at the ADJ similar to a "stick of rock" radiating out from the access (See Figure 7.1). After the threshold of material has been removed a popup message will be displayed asking the learner to consider the position of the ADJ in relation to cusps and fissures at any given point and to note their angulation. Later notifications will ask the learner to extend all sides of the access and introduce that caries at the ADJ should always be removed prior to caries at the cavity base. If at any point during the exercise an excessive amount of tissue is removed from the cavity base, a notification will draw the learner's attention to this and provide hints and reminders to monitor depth.

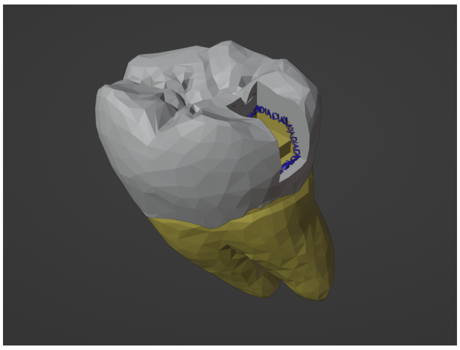


Figure 7.2: Exercise 1 Activity 2, Location of the ADJ Disto-Occlusally

Activity 1.2

Activity 1.2 is similar to activity 1.1, but here illustrates the ADJ when it is visible across the disto-occlusal surfaces. The concept is anchored by including an occlusal cut-away repeating content from activity 1.1 but then extending the accessed area to demonstrate how the ADJ continues to the distal aspect (See Figure 7.2). As above, the learner will be prompted to extend the access and explore the ADJ across these two surfaces. Notifications during tissue removal will draw attention to the vertical orientation of the ADJ and reinforce the concept that it is a *junction* between tissue-types rather than a planar concept.

Activity 1.3

The final activity of the first exercise demonstrates that access to and visibility of the ADJ is a dynamic concept during caries removal which can be impaired by the presence of unsupported enamel (See Figure 7.3). Introductory content will explain these concepts and instruct the learner to re-establish access to the ADJ by removing unsupported enamel until they have regained full visibility and access to the lettering. Once the threshold amount of unsupported enamel has been removed a notification will describe the importance of access to the ADJ when making operative decisions. The unsupported enamel will be configured as a target region, permitting the unsupported enamel to be highlighted in a different colour which can be toggled on/off.

Together, these 3 activities explore key concepts concerning the location of and what is meant by access to the ADJ. Breaking these concepts down in this way is intended to draw specific attention and encourage the development of low level schemas (Sweller et al., 1998) to establish component-part cognitive knowledge for later use in combination with other schemas developed in subsequent activities. The specific focus of each activity

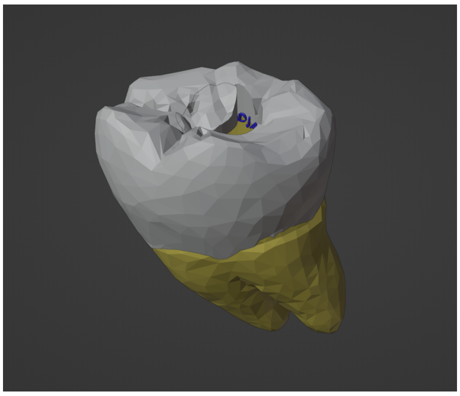


Figure 7.3: Exercise 1 Activity 3, Loss of access to the ADJ

is intended to manage the cognitive load of the concepts and give prominence to the intended information, reducing the risk that the learner becomes distracted by irrelevant information (Hammerton, 1967).

7.2.2 Exercise 2: Caries within the tooth and the infected to affected dentine transition

The activities of the second exercise focus on how caries spreads through the tooth producing varying degrees of demineralisation. It presents the language used to refer to caries in different regions of the tooth and introduces the idea that there is a prescribed order in which caries should be addressed. Activities will cover the following items from Table 7.1 and attention is focussed on the location of caries within the tooth and cues that indicate a change in the the extent of demineralisation:

- The need to manage pulpal floor caries separately and to address it separately as part of the corresponding goal
- The visual appearance of caries at the ADJ and the associated non-visual tactile sensations used when determining the extent of demineralisation
- The visual appearance and texture of the tissue being removed and how this relates to the level of demineralisation
- Criteria for selecting an appropriate size of bur for the removal of caries at the ADJ

Recognising the transition in demineralised tissue is a key area of attentional focus identified from the task analysis. Therefore, activities that promote the development of

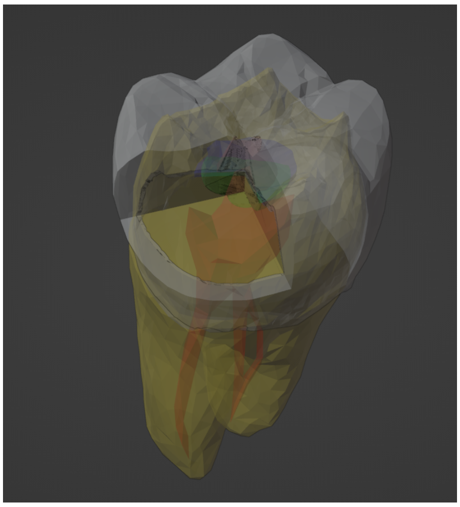


Figure 7.4: Exercise 2 Activity 1, Locations of occlusal caries

schemas to accomodate this knowledge and a confident awareness of this concept are likely to be useful for transfer activities in the future.

Activity 2.1

Activity 2.1 focusses exclusively on introducing the locations of caries in a lesion and the language used to refer to them. The exercise will be based around a partial cross-section of a tooth (See Figure 7.4), created by removing a cube of material encompassing approximately 25% of the crown. Brightly coloured regions will be placed within the tooth representing caries in the enamel, at the ADJ and deeper caries (including pulpal floor). Introductory text will inform the learner to commence the exercise by removing some of the carious enamel. After the threshold volume of tissue is removed, a notification will describe the region being interacted with, key features of caries in that region and explain its place in the order of operations. Because this activity is concerned with the location and language used to describe caries within the tooth, bright and unrealistic colours will be used to aid identification and more easily direct the learner's attention.

Activity 2.2

After establishing a shared language of where decay occurs, the learner's attention can now be directed towards the gradation of demineralisation. This activity displays a similar cut-away view as the previous activity, however, instead of a single colour, it shows concentric graduated bands illustrating different levels of demineralisation (See

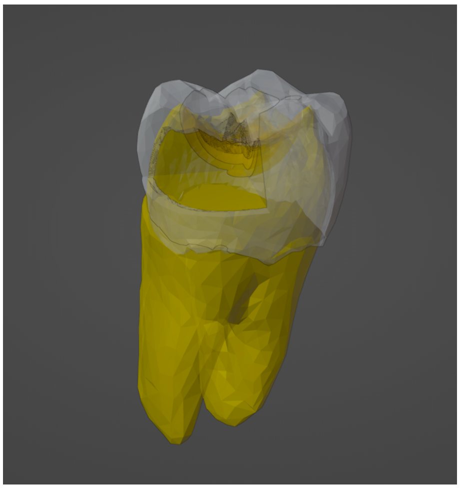


Figure 7.5: Exercise 2 Activity 2, Infected to affected demineralisation transition

Figure 7.5). Building upon earlier activities, learners will be directed to remove the caries at the ADJ, it will be pointed out that this is a simplified representation to explore the variation and the next exercise will introduce the structured sequence to its removal. If deeper caries is removed from the cavity base, a notice will be displayed to draw the learner's attention to this, re-stating the importance of addressing the ADJ first and reiterating guidance on where to focus attention to avoid the unintentional and premature removal of deeper caries.

As the learner gradually removes demineralised tissue, notifications will be displayed explicitly drawing their attention to the colour change, relating it to the real-world presentation, and to notice the tapering at the extent of the carious spread. As the learner approaches the extent, a notification will direct their attention to the size of the bur and its relation to the depth of caries being removed at this point.

These two activities are focussed on an exploration of the nature and location of caries within a tooth, taking steps to minimise additional cognitive load by introducing concepts one at a time. The next exercise builds on this further by introducing how different instruments are used methodically to establish a caries free margin at the ADJ.

7.2.3 Exercise 3: Clearing caries from the ADJ using a systematic approach

This exercise features a single activity which introduces the methodical approach to the removal of caries at the ADJ described during the task analysis. This exercise exercise integrates knowledge from previous content so much of the content from Row 3 of Table

7.1 is covered :

- The visual appearance of caries at the ADJ and the associated non-visual tactile sensations when determining the extent of demineralisation
- The appropriate depth of caries free margin to be cleared at the ADJ
- The presentation of sound tissue in order to determine that the limit of the caries has been reached and the ADJ has been successfully cleared
- The appropriate handpiece to use for each task within this goal
- How to recognise that enamel has become involved and is impairing visibility of the ADJ (Concept reinforced)
- How manage the removal of unsupported enamel and recognise when this has been achieved
- The motions used when removing caries at the ADJ with the low speed handpiece (Concept reinforced)
- The methodical one-internal-wall-at-a-time approach to manage the complexity of the task

Activity 3.1

The purpose of this activity is to guide the learner through the exploratory approach used when clearing caries from the ADJ and the instrumentation used at each stage. This activity is the most complex in terms of the number of addressable objects modelled and the context-specific support offered to learners whilst engaging with the exercise.

A quadrant cross section of a carious tooth will be presented (See Figure 7.6). Interaction with this exercise will be guided and learners will be prompted to follow the correct process via notifications displayed on screen following the successive removal of tissue. The caries will retain its banded-colours from the previous exercise to reinforce that the learner is removing caries, however the enamel will be displayed using distinctive colours to draw attention and assist with instruction:

1. Learners will be prompted to select a low speed handpiece and remove caries accessible at the ADJ. Once the threshold for removal has been reached attention will be drawn to the fact that the learner has lost access to the ADJ and created unsupported enamel (reinforcing content from earlier activities). Learners will then

be instructed to switch to the high speed handpiece[2] and carefully regain access to the ADJ.

2. After the unsupported enamel has been addressed, the learner will be informed that access has been restored and that they should, calling on knowledge from Exercise 2, inspect for the presence of caries at the ADJ and then re-select the low speed handpiece to address this newly accessible caries.

3. After a threshold amount of this caries has been removed, a notification will remind the learner to note that the severely infected tissue is yielding to affected. Once further material has been removed, a notification will indicate that access to the ADJ has been compromised again and that the unsupported enamel must be addressed to regain access.

4. During the 2nd iteration of unsupported enamel removal, a notification will draw the learner's attention to cues in the enamel that are associated with approaching the extent of the carious spread. When access to the ADJ is restored, a notification will ask the learner to assess the ADJ for the presence of caries again and repeat the cycle of material removal, re-selecting the low speed handpiece to continue. The notice will state that this alternating between hand pieces during caries removal is a core skill to be acquired.

5. By this stage only a small amount of carious tissue should remain. When a threshold amount of the total caries volume has been removed a notification will indicate that the visible caries has been addressed. The learner will be informed that not all demineralised tissue is discoloured so the a probe should be selected to assess this. The learner will be instructed to select the probe and then assess for any difference in tissue density at the ADJ. This step is exploiting an unsupported simulator feature as no difference between the tissues is detectable by the simulated probe. However, this act emphasises this step in the process so that it isn't diminished through omission.

7.2.4 Exercise 4: Reviewing and consolidating knowledge

Activity 4.1

The final exercise reinforces the *Evaluate ADJ* subgoal which completes the sub-steps of this goal. The learner's is reminded of the step's importance and then prompted to apply previous learning to the assessment of four quadrants of a pre-prepared cavity. The four quadrants will present as an exploded-view of the tooth as follows:

[2] NB only a single appropriate bur will be available for the high speed handpiece

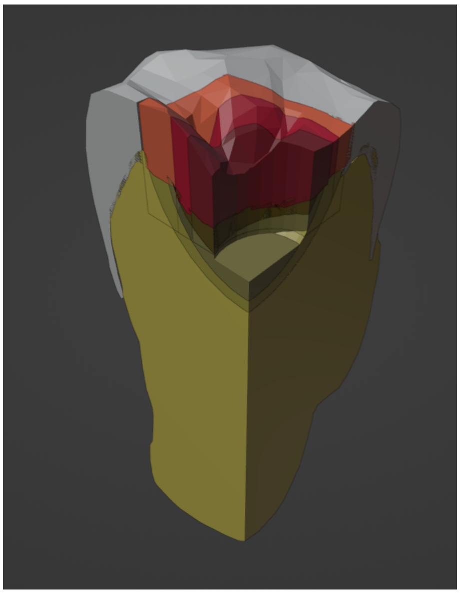


Figure 7.6: Exercise 3 Activity 1, Alternating between high and low speed handpieces

1. Quadrant 1 represents a preparation where caries at the ADJ has been correctly addressed (See Figure 7.7).
2. Quadrant 2 presents a preparation exhibiting an improperly cleared ADJ (See Figure 7.8).
3. Quadrant 3 presents a preparation where unsupported enamel has been retained (See Figure 7.9).
4. Finally, Quadrant 4 presents a preparation where pulpal floor caries has been prematurely addressed prior to establishing a caries free margin at the ADJ(See Figure 7.10).

Beside each quadrant, a yes/no disc of material will be displayed. The learner will be instructed to review each quadrant in turn and then cut into the corresponding disc if they feel that the quadrant has been adequately prepared or not.

For each quadrant where there are deficiencies in the preparation, if the learner answers correctly, a notification will acknowledge their answer and instruct them address the identified shortcomings. If the student answers incorrectly the notification will direct their attention to where the quadrant is deficient and instruct them to correct the shortcomings.

Collectively, across the four exercises, these activities introduce underlying concepts

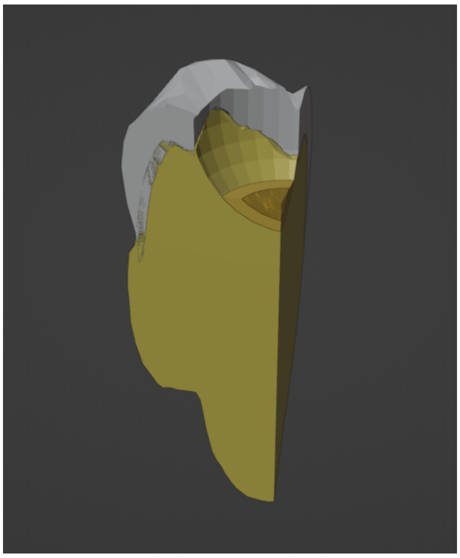


Figure 7.7: Exercise 3 Activity 2 Part 1, Correctly addressed quadrant

Figure 7.8: Exercise 3 Activity 1 Part 2, Improperly cleared ADJ quadrant

Figure 7.9: Exercise 3 Activity 1, Part 3 Unsupported enamel quadrant

Figure 7.10: Exercise 3 Activity 1, Part 4 Pulpal floor cleared before ADJ quadrant

key to developing an appreciation of establishing a caries free margin at the ADJ within a occlusal carious lesion. Each exercise focusses on a different and simplified aspect of the process in order to reduce the cognitive load and promote an increase in the effectiveness of the learning material.

The next chapter presents a Randomised Control Trial where these exercises are compared to traditional teaching approaches in the VR environment via a test of transfer.

Chapter 8

Comparing Deconstructed Learning with Traditional Teaching using VR Simulation

8.1 Background

In this chapter the deconstructed, part-task, activities focussing on the cognitive aspects of the goal of "Establish Caries Free Margin at ADJ" (presented in Chapter 7) are compared with the current whole-task approach to teaching. These part-task exercises eschew providing a whole tooth on which the learner can practice a whole procedure and, instead, focus on individual steps to provide "live" context-aware instructional messages to direct attention to relevant aspects of performance. Whole-task activities can exceed the processing capacity of a novice to perform them so part-task approaches are one way this complexity can be managed to reduce the overall time and effort required to acquire the skill (Paas and Merrienboer, 1994) which may improve the retention of the cognitive knowledge associated with the task.

This approach also connects to the theory of motor skill specificity which suggests that the reliability by which motor skills developed in a VR environment can be transferred to other contexts is limited because motor skills tend to be bound to the context in which they were acquired and are closely associated with the input signals received during training (Proteau, 1992). Any changes to these sensory inputs diminish the reliability of transfer and consequently improvements in performance as a result of the practice may only improve performance at the training task itself (Schmidt and Lee, 2014, pp. 218-219)

The above suggests that the current uses of VR in the dental curriculum may not be producing the desired improvements in performance. However, dental operative skills are a complex combination of motor performance and cognitive knowledge. Whilst the transfer of motor skill may be limited, manipulating the process by which cognitive knowledge is acquired should be encouraged to aid acquisition and transfer (Paas and Merriënboer, 1994). This presents an opportunity to explore alternative approaches using deconstructed part-task activities using the VR modality.

8.1.1 Aim

Determine if a series of cognitive-focussed part-task VR exercises result in enhanced transfer of operative knowledge when compared to existing whole-task VR training approaches.

8.1.2 Objectives

- Expose two equal groups of 1st year Dental Undergraduates to a VR educational intervention based on either part-task (experimental group) or whole-task (treatment-as-usual) approaches.
- Measure performance of participants on a series of transfer tests which:
  - Measure how many of the attributes of performance were retained from the material covered during the intervention session via a Retention Activity.
  - Measure the ability to apply the retained knowledge via a Ranking Activity
  - Measure the retention of procedural knowledge via a Procedural Activity
- Perform statistical analysis of the data to identify if a significant difference is present between the groups that can be attributed to the intervention.
- Capture participant opinion and explore themes relating to the face-validity and perception of the intervention via semi-structured interviews.

8.2 Method

To develop the method for this study, a full protocol was completed following the SPIRIT recommendations checklist (SPIRIT, 2013) as recommended by the BMJ as an evidence based consensus approach for interventional trials (BMJ, 2018). Whilst this chapter does not present a clinical study, it is not unreasonable that when exploring the acquisition of clinically relevant knowledge that a similar standard is aspired to. The methodology is summarised below and the full protocol for this study is presented in Appendix F.

This study is designed as a transfer test using a parallel group, two-arm, superiority trial with a 1:1 allocation ratio. The arms of the study are defined as an Experimental group and a Treatment-as-Usual (TAU) group.

Eligible participants will be randomised in equal numbers to the Experimental or TAU. Each group will undertake an intervention session consisting of taught content and exercises appropriate to their arm of the study followed by, after a cooling-off period, a 1:1 transfer test session (See Figure 8.1).

The intervention session for both arms of the study will take place in the simulation suite at the School of Clinical Dentistry, University of Sheffield, which contains 10 Virteasy (HRV, Laval) VR Dental Skills Trainers. The learning objectives of the intervention sessions will be aligned to the "Establish Caries Free Margin at ADJ" goal and tasks from the preceding task analysis study (See Chapter 5)

The 1:1 transfer session for both arms of the study will take place in an appropriate meeting room at the School of Clinical Dentistry. In this session participants will attempt to apply knowledge acquired during the intervention to a series of tests. The session will conclude with a semi-structured interview where participants explore themes around the acceptance and face-validity of the intervention.

8.2.1 Eligibility Criteria

Participants volunteering for the study must comply with the following at randomisation:

- Be enrolled in the 1st year of an undergraduate dental programme at the University of Sheffield
- Have completed the VR Suite familiarisation course.
- Must not hold an existing dental qualification (for example a qualified Dental Hygienist/Therapist or Dental Nurse seeking to extend their scope of practise).
- Must be on their first attempt (and not resitting) the 1st year of their respective programme.
- Be available to attend the intervention and transfer test session on the agreed dates.

The timing of the study ensures that participants are not undertaking any content in their degree programme that would confound the goals of this study.

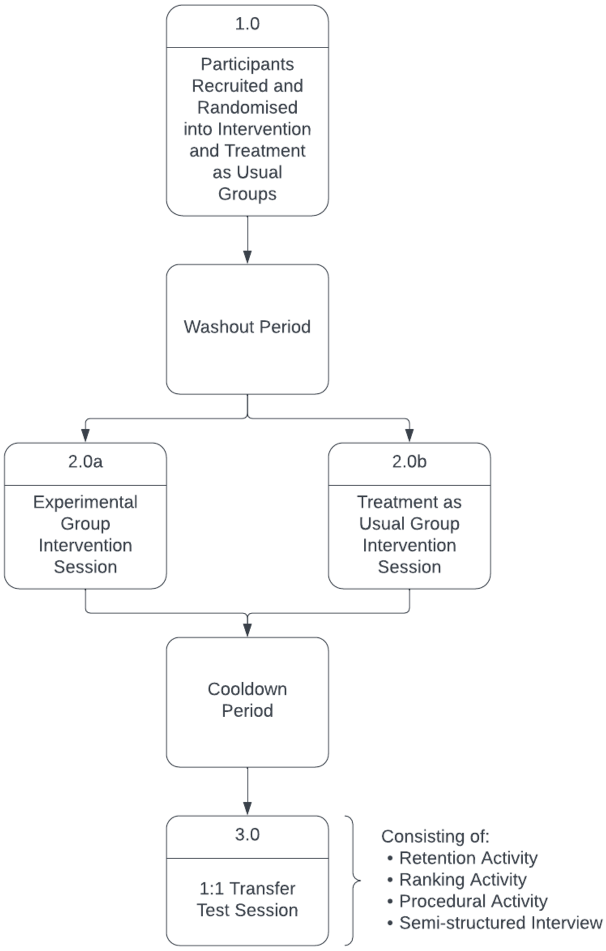


Figure 8.1: Flow Chart of Comparative Study Method

8.2.2 Intervention Session

Treatment-as-usual Group

Participants assigned to the treatment-as-usual group will undertake a 1-hour training session in the simulation suite. This will commence with a tutor-led presentation (with supporting slides) to introduce the procedure for caries removal at the ADJ followed by the use of existing simulation exercises to apply the content of the presentation. The learning objectives and material covered in the Treatment-as-usual session will be equivalent to that of the experimental group but will use whole-task simulation exercises instead of the focussed part-task exercises used by the experimental group.

Participants will be free to ask questions and seek clarifications. A note will be made of the participant and the question asked for later reference.

Experimental Group

Participants assigned to the experimental group will undertake a 1-hour training session in the simulation suite. The session will commence with the same introduction to the task followed by interaction with the part-task exercises developed for this study. Each of the exercises in this series present an activity aligned to acquiring knowledge which contributes to the learning outcomes.

Participants will be free to ask questions and seek clarifications. A note will be made of the participant and the question asked for later reference.

8.2.3 Transfer Test Session

After a 1-week cool down period, participants will attend a transfer test session where they will undertake a series of exercises to measure the effectiveness of various aspects of the intervention. The transfer session will be recorded (audio and video capturing the participants hands only). After the conclusion of the session, the recording will be reviewed to extract data relating to performance for analysis. Both arms of the study will undertake the same transfer test activities. All tests are designed to test knowledge related to the "Establish Caries Free Margin at ADJ" goal from the task analysis.

Retention Activity

The aim of this activity is to measure the participants retention of declarative facts. Participants will be presented with (n=5) 3D printed tooth models. Each of these models will contain a carious lesion which has been prepared in anticipation of a restoration being placed. Participants will be asked to take each tooth in turn and verbally identify attributes of performance based on the state of the preparation, drawing upon the content covered during the intervention session. Participants will be provided with access to a dental probe to assist them in their evaluations.

Statistical tests for the retention activity will explore difference in means between the two arms of the study in terms of total number of attributes retained, the number of times each attribute was identified across multiple models, the number of mis-identifications of incorrect attributes and a negatively-marked Retention Score.

Ranking Activity

The aim of this activity is to assess the participants judgement and application of the retained knowledge. Participants will be asked to rank the five teeth that they have just critiqued into order of clinical acceptability. The most appropriately prepared tooth is the one they consider to be most suitable for immediately progressing to the next stage of treatment, the least appropriate is the one they consider having the most egregious issues. Once they have completed the activity, they will be asked to explain the reasoning they used.

Statistical tests will be performed for each arm of the study to establish the mean ranking of each model, a pairwise comparison and a Kendalls Coefficient of Concordance. Next, each participant's rankings will be compared to a rational scoring key (provided by an expert) and the sum of the distance-squared calculated to produce a 'score' for each participant. A difference in means test will be performed on these scores to identify if a statistical difference exists which might be attributed to the intervention received.

Procedural Activity

The aim of this activity is to assess the participants retention of the procedural steps required to establish a caries-free margin at the ADJ. Participants will be provided with a tooth model (different from those used in the previous activities) with a partially-prepared cavity and will be asked to describe which steps they would take (and the instruments they would use) in order to complete the goal.

Performance on the procedural activity will be scored by an assessor, blinded to the intervention, from the recording using a standardised rubric marking scheme derived from the task description. A test for a difference in means will be performed on the scores awarded to identify if a statistical difference exists between the two arms of the study.

Semi-structured Interview

The transfer test session will conclude with a semi-structured interview to capture participant opinions on the intervention received. Face validity is an important concept in educational approaches and can impact upon engagement with the material. Responses to these questions will be thematically analysed using the approach presented in Braun and Clarke (2006) in order to identify any common themes in the perceptions of the approach. Photographs and a description of the experimental approach will be provided to participants assigned to the TAU group to allow them to share their initial impressions.

8.3 Results

Following ethical approval (University of Sheffield Ethics Reference: 058225), an invitation to participate in this research was circulated via email to 1st year students enrolled on the Bachelor of Dental Surgery programme at The University of Sheffield on the 15th of January 2024. Participants were further reminded of the invitation and given an opportunity to ask informal questions at the end of scheduled teaching sessions held during the period 22nd of January 2024 to the 2nd of February 2024. These recruitment efforts resulted in 31 responses being submitted to the online intention to participate form. Of the 31 responses, 4 were rejected: 3 being identified as duplicate submissions and 1 was submitted by a volunteer who did not meet the eligibility criteria of being on their first sitting of the first year of the programme.

The 27 eligible participants were randomised into the two arms of the study using a computer-generated randomisation schedule and sent a calendar invitation to an intervention session during which formal consent was recorded. Intervention sessions took place following a minimum of a 3 week washout period after the participant's last timetabled session in the VR Simulation suite. One volunteer did not attend their allocated intervention session and did not respond to emailed offers to re-schedule to another group so was excluded from the study.

The 26 participants who attended the intervention session also attended the transfer test session and successfully completed the study. A target of 27 participants was set as an achievable recruitment target for this study based on the available population. A sample of 27 participants would achieve a confidence level of 95% with a confidence interval (margin of error) of 15. Recalculating the confidence interval with the sample

Figure 8.2: Transfer Test Session Setup

of 26 completed participants returns a slightly reduced confidence interval of 15.5 at the 95% confidence level. Full details of the sample size calculation are provided in Appendix F.

The protocol planned a 7 day interval between the intervention and the transfer test, however, due to participant availability this could not be achieved in all instances. The mean interval between intervention and transfer session was 7.67 days with a mode of 7 days. The maximum interval was 12 days and the minimum was 17:00 on the 6th day which was rescheduled at the participant's request from 09:00 on the 7th day following the intervention. Between the intervention and transfer test all participants confirmed that they had not undertaken any further study (independently or taught) related to caries removal.

The following sub-sections will present statistical calculations performed on, and visualisations of, participant responses during the transfer test (Figure 8.2) to the activities exploring individual performance and inter-group differences that may be attributed to the intervention received. Source data that these calculations are based upon is provided in Appendix J.

The primary outcome measures for this study were all quantitative in nature. Participant performance was assessed following the intervention using tests that explored the retention of declarative facts, application and reasoning about the content and the retention of procedural knowledge.

To compile the data for analysis, video recordings of the session were reviewed and data captured using a data extraction form (see Appendix H). Data from the completed forms were transcribed into an Excel (Microsoft, 2024) spreadsheet for subsequent processing using an R (R Core Team, 2023) script. The following sections describe the results derived from the execution of the script; the source code for which is listed in Appendix G.

8.3.1 Retention Activity

The retention activity asked participants to identify attributes of performance present in 3D printed models of carious/prepared teeth (n=5). All models were printed using a Stratasys J5 DentaJet (Minnesota, USA) with patterns of caries painted by hand using Citadel Colour (Games Workshop, Nottingham, UK) acrylic paints (Figure 8.3). Results were compared between the experimental and treatment as usual (TAU) group to identify any differences in performance that could be attributed to the intervention received. The tooth models were presented to the participant in a randomised order, asking them to recall information retained from their intervention session to identify the presence of positive or negative attributes of performance present in each model. Performance between groups was compared along the following dimensions:

- Total Correct: the sum of the correctly identified attributes. Superiority is defined as the greatest number of performance attributes correctly identified
- Total Incorrect: the sum of attributes incorrectly identified. Superiority is the fewest number of attributes misidentified
- Total Retention: a raw comparison of whether the attribute was retained and applied, regardless of if it was applied correctly or not. Superiority is defined as the greatest number of attributes retained.
- Retention Score: a negatively marked indicator calculated as total correct minus total incorrect. Superiority is the highest score.

Finally, individual models were explored to establish if either intervention biased performance when demonstrating the acquired skill on any particular model.

Figure 8.3: 3D printed models used for comparative study retention and ranking activities

Total Correct

As a non-negative integer count of independent events occurring within a given time frame, the count of correctly identified attributes is hypothesised to follow a Poisson distribution. Establishing this is the case will permit appropriate tests to be performed to measure if there is a significant difference between the two arms of the study. To test that the results follow a Poisson distribution, a Goodness of Fit test was performed using a minimum chi-square test to evaluate if the data significantly differed from the Poisson distribution. The relationship between these variables was not significant, χ^2 (19, N = 26) = 11.928, p = 0.8886832. As P>0.05 the alternative hypothesis that the data is from two populations is rejected and the data can be assumed to follow a Poisson distribution.

The experimental group achieved a mean total correct attributes identified of 13.92 with a range of [9-18]. The TAU group achieved a mean total correct attributes of 13.15 from a range of [7-20]. The Poisson Rate test indicates that the rate of correct identification of attributes has a value of 1.058 with a p-value of 0.6315 and therefore the difference between the two groups is not statistically significant (P>0.05) and the alternative hypothesis can be rejected.

Total Incorrect

Following the same rationale as above, the count of Total Incorrect is hypothesised to follow a Poisson distribution and a minimum chi-square test was again used to evaluate if the data significantly differed from the Poisson distribution. The relationship between these variables was not significant, χ^2 (10, N = 26) = 37.242, p = 5.138525e-05. As P>0.05 the alternative hypothesis that the data is from two populations is rejected and the data can be assumed to follow a Poisson distribution.

The experimental group misidentified a mean average of 3.38 attributes with a range of [1-10]. The TAU group misidentified a mean average of 5.23 attributes from a range of [1-11]. The Poisson Rate test for these results returns a rate ratio of 0.647 and a p-value of 0.0293. As P<0.05, this indicates that the difference between the two arms of the study is statistically significant and the null hypothesis can be rejected. In this study, the experimental group made fewer mis-identification errors than the TAU group.

Overall Retention

Correctly identifying attributes of performance is always preferable, however, recall is a precursor to application so simply recalling the existence of attribute is of value even if it is subsequently applied incorrectly. Therefore, the incorrect and correctly identified attributes were combined to explore the presence of a difference in the raw retention of the attributes between the arms of the study. As both sets of results have been shown to be from the Poisson distribution, then the union of the two sets is also from the Poisson distribution.

The experimental group achieved a mean overall retention score of 17.31 attributes with a range of [13-21]. The TAU group identified a mean of 18.38 attributes with a range of [13-23]. The Poisson rate test returned a rate ratio of 0.9414226 and a p-value of 0.5462 so this difference is not statistically significant and the alternative hypothesis can be rejected.

Negativly-marked Retention Score

Whilst recall is a precursor to application, the guessing of answers is a negative indicator in a professionally accredited course that values practitioners being reflective and aware of the limits of their knowledge (GDC, 2015). Therefore, a score can be derived by subtracting the number of incorrect answers from the number of correct responses, providing a negatively marked retention score. Using this measure, the Experimental group received a mean score of 10.538 from a range [1-17], whereas the TAU group received a score of 7.923 from a range of [1-19]. This difference was found to be statistically significant ($p < 0.05$) with a Poisson Rate Ratio of 1.330097 and a p-value of 0.03294 therefore, the null hypothesis that there is no difference between groups is rejected.

Exploring Bias in the Individual Models

It is possible that demonstrating the acquired skill could be biassed by attributes of individual models and lead to unreliable measurements. Therefore, it is necessary to compare the correctly and incorrectly identified attributes ascribed to each individual model across the two arms of the study. Tables 8.1 and 8.2 list the total number of correct/incorrect attributes available for identification (respectively) followed by the mean average of those attributes identified by the experimental and TAU groups. The Poisson Rate ratio and p-value for performance by the two arms of the study are also calculated.

No statistically significant difference could be found in the identification of correct or incorrect attributes between participants of either group. Therefore, the alternative hypothesis that performance at retention tests for each model are impacted by the intervention received is rejected.

Beyond the counts of correctly and incorrectly identified attributes, it is also informative to compare the standard deviations across responses. For the incorrect attributes, the responses from the Experimental Group had a standard deviation of 0.831 and the

	Available Correct Per Model	Experimental Group Mean	TAU Group Mean	Poisson Rate Ratio	p-value
Diamond	8	2.538	3.231	0.786	0.3557
Triangle	8	2.154	2	1.077	0.8919
Square	8	3.692	2.692	1.371	0.1875
Circle	9	2.615	2.692	0.971	1
Rectangle	8	2.923	2.538	1.151	0.6353

Table 8.1: Average correct attributes identified by participants and significance of difference

	Available Incorrect Per Model	Experimental Group Mean	TAU Group Mean	Poisson Rate Ratio	p-value
Diamond	8	0.769	0.846	0.909	1
Triangle	8	1	1.615	0.619	0.2295
Square	7	0.308	0.846	0.364	0.1185
Circle	9	0.692	1.385	0.5	0.1221
Rectangle	8	0.615	0.538	1.143	1

Table 8.2: Average incorrect attributes identified by participants and significance of difference

TAU Group had a Standard Deviation of 0.975. The F Test was used to compare the standard deviation and showed a ratio of variances of 0.726 with a p-value of 0.2039, therefore the differences between the two standard deviations is not statistically significant. Performing the same calculations for the correctly identified attribute shows a standard deviation of 1.244 for the experimental group and 1.054 for the TAU group. The F Test shows a ratio of variances of 1.391 and a p-value of 0.1891. Again, showing the differences between the two standard deviations are not statistically significant and the alternative hypothesis be rejected.

Comparison of attributes across models:

Having established no evidence of bias in the models, performance can be aggregated across them to explore if either arm of the study is more likely to have retained and be able to recall a particular attribute of performance when it occur across multiple models. This is a measure how many times the attribute is recognised or omitted in total and if there is a difference between the two arms of the study. Five attributes are present across multiple models and are suitable for aggregation in this way (See Table 8.3). Because the identification of the attributes (or failure to identify) is a binary proportion, the significance test for this comparison uses a Pearson's Chi-squared test with

	Experimental Group		TAU Group		Pearson's χ^2
	Identified	Omitted	Identified	Omitted	
Identification of caries at ADJ	53	12	52	13	$\chi^2 = 0$, p-value = 1
Identification of unsupported enamel	18	47	22	43	$\chi^2 = 0.325$, p-value = 0.5686
Recognition of retained caries to pulpal floor	12	40	14	38	$\chi^2 = 0.051282$, p-value = 0.8208
Appraisal of cavity form	8	57	10	55	$\chi^2 = 0.064484$, p-value = 0.7995
Use of probe to assist assessment of cavity	37	15	26	26	$\chi^2 = 4.0263$, p-value = 0.0448

Table 8.3: Retention of attributes present on multiple models

Yates' continuity correction. Results for this comparison across each of the attributes of performance is presented in Table 8.3.

No statistically significant difference was identified across four of the attributes of performance: Identification of caries at ADJ, Identification of unsupported Enamel, Recognition of retained caries to pulpal floor or Appraisal of cavity form. Therefore, for these attributes the alternative hypothesis can be rejected. However, the attribute of "Used probe to assist assessment of cavity" has a p-value of 0.0448. As $P < 0.05$, this indicates a statistically significant difference between the two arms of the study. Therefore, in this study, the experimental group was more likely to use a dental probe when assessing a cavity than the TAU group and the null hypothesis can be rejected.

During data extraction, it was identified that participants occasionally truncated their responses and may not have fully demonstrated their ability. For example, some participants were observed to describe in detail both positive and negative attributes of a deficient preparation, but when presented with a subsequent model that was more acceptable simply described it as "ok" or "that one looks good". Clearly, having just demonstrated an awareness of a range of relevant attributes which were retained from the intervention, it is possible that they elected to not immediately repeat them for brevity or simply assumed there was an implication based on their recent comments. As the participants were blinded to the assessment method, they would not have known that repeating the attribute was desirable but prompting them may have introduced bias into the responses by hinting that they had missed something. To address this observed phenomena, data imputation was applied to the responses. Data imputation provides a method to systematically "fill in the blanks" and, where appropriate, allow a participant to be given the benefit of the doubt. To impute the omitted data a rule using single imputation was applied whereby if a participant had correctly identified an attribute in at least two other instances and did not ever misidentify that attribute then it was assumed the participant understood the attribute and would have corrrectly described it if prompted. The rule was applied to all data, however, this step did not change the outcome of any statistical analysis therefore these results are not included here.

8.3.2 Ranking Activity

After identifying attributes of performance for each model, participants were asked to rank the same models in the order of which they considered best to worst; the best being the model most suitable to immediately proceed to the "Address Caries at the Pulpal Floor" stage. Participant rankings were first assessed using methods to explore the level of agreement and consistency of ranking within each arm of the study. Next, the participant rankings were compared to a rational key taken from an expert opinion of the most clinically correct ranking (see Table 8.4). The difference between the participant and expert ranking was used to derive a score by which performance in this task could be compared between arms of the study.

The remainder of this section will detail the measurements and comparisons performed on the results of the Ranking Activity.

Group Agreement

Kendall's Coefficient of Concordance, or Kendall's W, is a measure of agreement between a set of rankings. A higher score (measured between 0 and 1) indicates that there is more agreement within the group, with a lower score suggesting more disagreement between individual rankings. The rankings of the Experimental Group achieved a Kendall's W of: 0.166 with a p-value of 0.07146588 and the TAU group achieved a Kendall's W of 0.196 with a p value of 0.03695171. This indicates a low agreement between members of both groups. The level of agreement was statistically significant for the TAU group ($p < 0.05$) but not statistically significant ($p > 0.05$) for the Experimental group.

A ranking task can be investigated via descriptive statistics to explore the relationships between the rankings and the tendencies of each arm of the study:

	Key Rank	Comments/Rank Rationale
Diamond	1	This model was correctly prepared and shows no deficiencies relevant to this stage of the overall procedure
Triangle	2	This model is similar to Diamond above, however, a region of unsupported enamel prevents confirmation that a caries free margin has been fully established
Square	3	Caries can be seen at the ADJ distally. Additionally, angulation of the bur was poor resulting in bulging at the base of the cavity. Depth control was poor and some areas show premature removal of caries at the Pulpal Floor.
Circle	4	Caries can be seen at the ADJ mesio-lingual, mesial, mesio-buccal and distally. Pulpal floor has been completely cleared of caries; no affected tissue retained.
Rectangle	5	Caries free margin established at ADJ and no unsupported enamel present, however, the tooth is excessively over-prepared. Unlike the above models where further effort can produce a satisfactory preparation, this can no longer be achieved due to the excessive tissue removal.

Table 8.4: Subject matter expert Rational Key ranking and rationale (best to worst)

	Experimental Group		TAU Group	
	Mean Rank	Standard Deviation	Mean Rank	Standard Deviation
Diamond	1.92	1.754	2.31	1.797
Triangle	3.15	1.573	2.31	1.437
Square	3	1	3	1.354
Circle	3.31	1.182	3.54	0.877
Rectangle	3.62	1.044	3.85	0.899

Table 8.5: Average Ranking and Standard Deviation for each model

	Diamond	Triangle	Square	Circle	Rectangle
Diamond		12	12	12	10
Triangle	1		11	8	6
Square	1	2		8	4
Circle	1	5	5		3
Rectangle	3	7	9	10	

Table 8.6: Experimental Group Pairwise Comparison

Average Ranking

Calculating the average ranking allows comparisons between the average rank provided by each arm of the study and further comparisons to the rational key. Mean average rankings and standard deviations for each arm of the study are presented in Table 8.5.

Pairwise Comparison

Further exploration of the rankings performed by each group can be carried out through a comparison of the pairwise rankings between the models. This provides a representation of how many times a given model was ranked higher than another. Pairwise comparisons for each model are shown for the Experimental group in Table 8.6 and for the TAU group in Table 8.7. This table should be read row-to-column so for example in Table 8.6, Triangle is ranked higher than Rectangle 6 times.

	Diamond	Triangle	Square	Circle	Rectangle
Diamond		12	12	12	10
Triangle	1		11	10	7
Square	1	2		8	5
Circle	1	3	5		4
Rectangle	3	6	8	9	

Table 8.7: Treatment As Usual Group Pairwise Comparison

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Diamond	10	2	0	0	1
Triangle	0	4	5	4	0
Square	0	1	4	4	4
Circle	0	2	3	2	6
Rectangle	3	4	1	3	2

Table 8.8: Marginal Matrix for Experimental group

	1^{st}	2^{nd}	3^{rd}	4^{th}	5^{th}
Diamond	8	4	1	0	0
Triangle	0	6	5	1	1
Square	1	0	3	6	3
Circle	1	1	1	4	6
Rectangle	3	2	3	2	3

Table 8.9: Marginal Matrix for Treatment As Usual group

Finally, a marginals matrix presents number of times an item is ranked in a given position. The marginals matrix for the Experimental group is presented in Table 8.8 and for the TAU group in Table 8.9.

Comparison to Rational Key

The above analysis is done without reference to any notion of "correctness" with regard to the rankings suggested by the participants. To assess if either arm of the study returned rankings that were closer to the rankings submitted by a subject matter expert, a score for each participant's ranking was calculated. This score was taken from the sum of the square difference in the participants ranking to that of the ranking suggested in the rational key. For example, if a participant ranked a model in 2^{nd} place and the expert ranked that model in 4^{th} place, then the difference between these rankings is 2; $2^2 = 4$ which would be added to the participant's total score. Superiority for this scoring is defined as the lowest number where a ranking identical to the rational key is equal to a score of 0.

Using this scoring mechanism, the mean average score for the Experimental group was 12.923 from a range [0-40] and the TAU Group was 11.385 from a range of [0-26], suggesting that the TAU Group outperformed the Experimental Group on this metric. However, an outlier is present in the Experimental group where a participant suggested a ranking the exact opposite of the expert ranking and consequently received a score of over 3x higher (worse) than the group average. This outlier may be skewing the scores; removing this value results in an average Experimental Group score of $10.6\dot{6}$ from a range of [0-22] which now suggests the Experimental Group outperformed the TAU Group.

To explore the significance of these results, the scores were first assessed to establish if they followed a normal distribution. Firstly, by viewing a Quartile-Quartile plot (Figure 8.4), which suggested that the values lie within the normal distribution (indicated by the shaded area) and then confirmed with a Shapiro-Wilk test which did not show a significant departure from normality, W(26) = 0.93, p-value = 0.059.

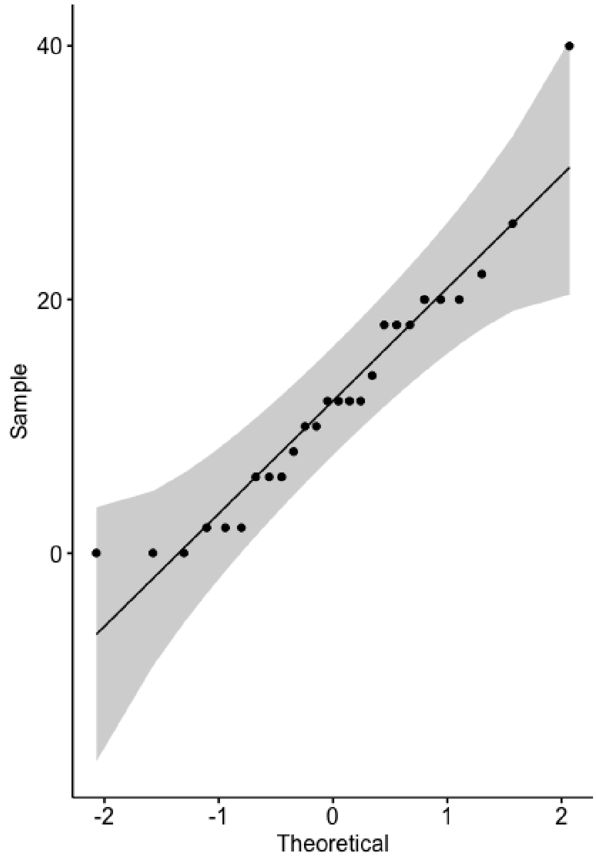


Figure 8.4: Rational Key - Normal Distribution of Scores

A normal distribution permitted the use of a Two Sample t-test, which returned t = 0.407, df = 24, p-value = 0.6878. Therefore, the difference between the two samples is not statistically significant ($p > 0.05$) and the null hypothesis cannot be rejected.

Excluding the outlier noted above has no effect on these results: the Shapiro-Wilk test does not show a significant departure from normality in the sample (W(25) = 0.94, p-value = 0.1746) and the Two Sample t-test does not show a statistically significant difference (p-value = 0.8226).

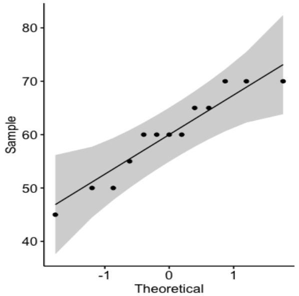
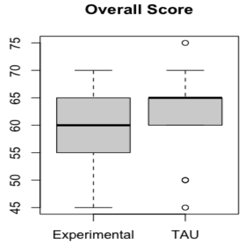
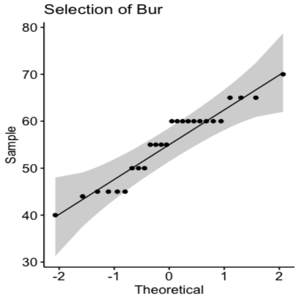
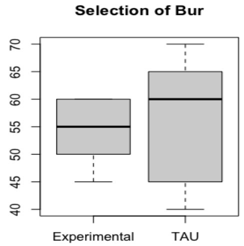
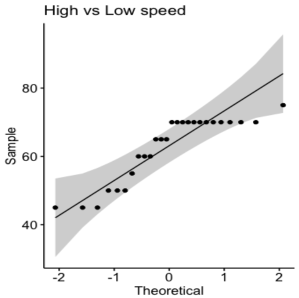
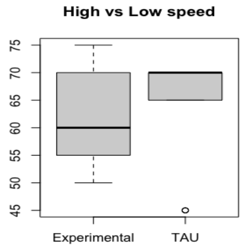
Figure 8.5: 3D printed model used for comparative study procedural activity

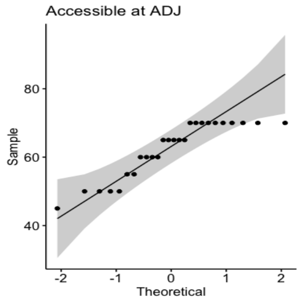
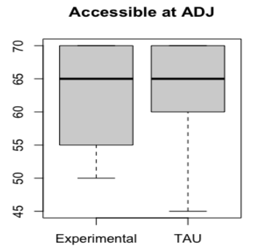
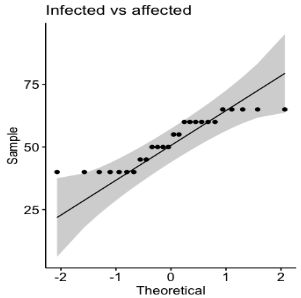
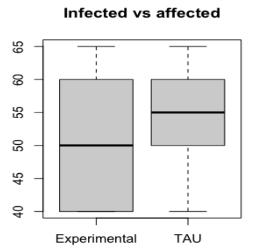
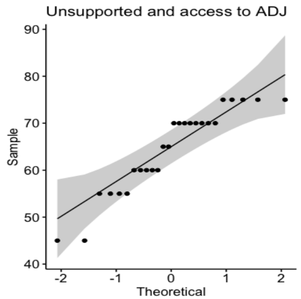
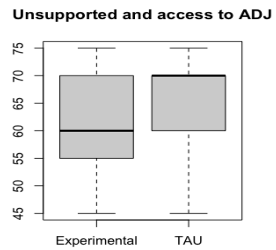
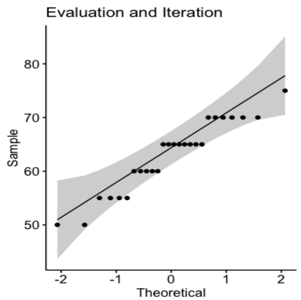
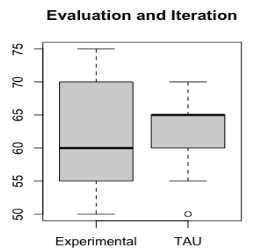
8.3.3 Procedural Activity

The Procedural Activity assesses the participant's ability to recall the steps, in order, and any relevant considerations to achieve the goal of "Establishing a Caries free margin at the ADJ". Participant performance for this task was assessed by reviewing a video recording of the participant verbally describing what steps they would take and which instrumentation they would use to achieve the goal state from a starting point presented in a 3D printed model of an accessed carious lesion (Figure 8.5). After reviewing the video recording, participant performance was assessed against a rubric consisting of 7 separate criteria (See Appendix I). These criteria aligned to the steps expected to successfully complete the goal plus an overall mark. For each assessed element, a percentage score was awarded.

Analysis was performed on each of the criteria separately and on the overall awarded score. Data for each score were tested for normality (both visually via a Quartile-Quartile plot and statistically by applying a Shapiro-Wilk test to the data) and then an appropriate test statistic was calculated to establish if the difference between the two groups was statistically significant. Where the data was found to be normally distributed, a Two Sample t-test was performed and where non-normally distributed data was found, a non-parametric Mann-Whitney test was performed. Finally, to visualise the data, box plots were produced to illustrate the distribution of marks across each of the arms of the study.

The results of the above statistical tests and associated charts are shown in Table 8.10. No statistically significant differences were detected between the two groups across any of the measurements of the procedural activity.

Assessment	Quartile-Quartile Chart	Shapiro Test	Statistic	Standard Deviation and F Test	Box Plot
Overall Score		W(26) = 0.920, p-value = 0.2484	Two Sample t-test: t = -0.693, df = 24, p-value = 0.4951	Experimental: 6.504, TAU: 5.718, F = 1.294, num df = 12, denom df = 12, p-value = 0.6623	
Selection of Appropriate Bur		W = 0.928, p-value = 0.06804	Two Sample t-test: t = -0.711, df = 24, p-value = 0.4839	Experimental: 5.718, TAU: 9.761, F = 0.343, num df = 12, denom df = 12, p-value = 0.07596	
Use of high and low speed handpiece		W = 0.813, p-value = 0.0002922	Mann-Whitney U test: W = 81.5, p-value = 0.8923	Experimental: 9.041, TAU: 10.516, F = 0.739, num df = 12, denom df = 12, p-value = 0.6088	

Removal of tissue at ADJ		W = 0.841, p-value = 0.0009461	Mann-Whitney U test: W = 73, p-value = 0.5585	Experimental: 8.006, TAU: 8.549, F = 0.877, num df = 12, denom df = 12, p-value = 0.8242	
Distinguishing between Infected/Affected		W = 0.859, p-value = 0.002129	Mann-Whitney U test: W = 67, p-value = 0.3731	Experimental: 10.377, TAU: 9.094, F = 1.302, num df = 12, denom df = 12, p-value = 0.6545	
Management of unsupported enamel/regaining access to ADJ		W = 0.892, p-value = 0.01047	Mann-Whitney U test: W = 66, p-value = 0.3445	Experimental: 9.707, TAU: 8.204, F = 1.4, num df = 12, denom df = 12, p-value = 0.569	
Evaluation/Inspection of ADJ and iterative nature of task		W = 0.931, p-value = 0.08075	Two Sample t-test: t = -0.435, df = 24, p-value = 0.6678	Experimental: 7.532, TAU: 5.911, F = 1.624, num df = 12, denom df = 12, p-value = 0.4131	

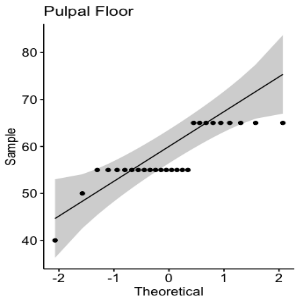
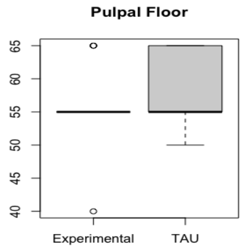
Removal of caries from pulpal floor		W = 0.755, p-value = 3.269e-05	Two Sample t-test: t = -1.281, df = 24, p-value = 0.2124	Experimental: 6.504, TAU: 5.718, F = 1.294, num df = 12, denom df = 12, p-value = 0.6623	

Table 8.10: Procedural Activity recall scores

8.3.4 Participant Perception

Secondary outcome measures were intended to explore participant perceptions of the study and identify themes from semi-structured interviews. The interviews were conducted at the end of the transfer session and asked participants questions relating to the following topics:

- Their general thoughts on the exercises used during their intervention session and how they could be improved
- Thoughts on how the Experimental/TAU exercises compared and their preferences (TAU group participants were shown a brief overview of the experimental approach and asked to share their initial reactions)
- The realism of the simulation and its importance
- If they felt prepared for the Transfer Test

Additionally, it was intended that the experimental intervention would use software pop-up dialogues based on user input to inform and guide their actions. These were not successfully implemented in time for the Intervention sessions, therefore this idea was presented and discussed with participants during the interview to explore their thoughts on such a feature being added in the future.

Thematic Analysis

The interviews were analysed using thematic analysis following the approach laid out in Braun and Clarke (2006). Familiarity with the interviews was established by transcribing participant comments (including annotations for non-verbal gestures and clarifications from the interviewer). These transcripts were then reviewed and initial codings annotated against the narrative. These initial codings were reviewed and refined then themes and sub-themes identified which could be explored with reference to relevant literature to draw insight into the efficacy and appeal of the approach as perceived by the participants. The analysis follows a theoretical thematic analysis approach (Braun and Clarke, 2006), so whilst themes are linked to the data itself (and where relevant prevalence is stated), the analysis is rooted in the skills development literature. This allows insightful perspectives to be discussed even if they were only raised by a small number of participants rather than being restricted by a measure of prevalence. At the time of the interview participants had no indication of their performance on the transfer task so Knowledge of Results (Schmidt and Lee, 2014, p. 258) would not have influenced their comments during the interview.

The themes identified from the interviews are as follows:

- **Attentional Focus:** This theme explores how the intervention directed participants attention to relevant attributes and how this contributed to their learning
- **Preparing for practice:** Explores how the intervention prepared participants for the transfer test and their thoughts on how these exercises might prepare them for transition to clinic later in the course
- **Preferences for learning and skills development:** This theme explores participant opinions on the intervention received and how it supported their skills development
  - **VR as a teaching tool:** This sub-theme explores opinions on the use of VR as a teaching tool and their thoughts on the introduction of pop-up notifications as a teaching aid.
- **Intervention delivery:** Finally, participants framed the intervention as a taught session and provided comment on which elements they liked or disliked and how these might influence to future session design.

Edited participant transcriptions and associated analysis tagging from the interviews are available in Appendix K.

8.4 Discussion

8.4.1 Retention Activity

Across the 4 measures of retention (See Section 8.3.1), the Experimental group outperformed the TAU group on 3 of the measures: Total Correct, Total Incorrect and the negatively-marked Retention Score. The TAU group achieved a higher score for the Total Retention measure. Of these measures the scores for Total Incorrect and the Retention Score achieved statistical significance ($p < 0.05$).

At this early stage in their training, recalling an attribute *at all* regardless whether it was applied correctly is notable. In this measure of Total Retention the TAU group outperformed the Experimental group, however, this did not achieve statistical significance ($p > 0.05$). This measure, however, is vulnerable to being 'gamed' through indiscriminate listing of attributes. Whilst there is no suspicion that participants maliciously did so, this weakness rewards a participant who lists every attribute they can recall and leads to a score which is not representative of their true ability. Furthermore, it could be argued that the inverse is also true and the measure penalises an understanding of the attributes because the participant who *knows* that an attribute is not present would not mention it at all and therefore not gain a 'point'. This weakness gives credit to incorrect answers and does not recognise when they have been left unsaid because the participant correctly applied the cognitive knowledge associated with that term so it is not a sound basis to drawing strong conclusions in isolation. Therefore this test has a limited validity because a learner can perform well on the test without understanding the relevant theoretical concepts (Messick, 1995) permitting performance to be artificially high due to lack of penalty attached to other aspects of the test (Chambers, 2012).

Separating this measure of Total Retention into its constituent parts of correctly identified and misidentified attributes provides some protection against the above weaknesses. The Experimental group identified slightly more (but not significantly) attributes correctly but also made significantly ($p < 0.05$) fewer errors of misidentification. The Total Correct score is still vulnerable to being 'gamed' in the same way as the Total Retained, however, the measure of Total Incorrect is resilient to this as it negatively marks the use of this strategy by penalising inaccurate responses. The participants were blinded to the assessment measure, but, considering these results in combination suggests that the TAU group may have a weaker understanding of the material and were recalling attributes without regard to their correctness. That the experimental group made statistically significantly fewer errors would suggest that this group were able to demonstrate a greater degree of discrimination when identifying attributes of performance in the models.

These two measures can be further analysed by subtracting the number of incorrect attributes identified from the total correctly identified to derive a negatively-marked score of retention. This measure shows that the Experimental group statistically significantly ($p < 0.05$) outperform the TAU group. The use of negative marking is arguably crude as it implies an equivalence and commutability of correct and incorrect responses whereby a minor or inconsequential incorrect response negates an important correct answer, however, as a mechanism to balance a demonstration of underlying knowledge against a demonstration of recall it is considered acceptable as a limited measure through which to explore performance across the two arms of the study.

Some attributes of performance were present across multiple models and these allow for retention to be explored in aggregate. In most cases, no statistically significant differences were detected between the two arms of the study, however, one exception to this was the measurement of the use of the dental probe to assist in the assessment of the cavity. The use of the probe is perhaps a more procedural aspect of carrying out the activity (rather than the more declarative nature of identification of the presence/absence of attributes), however, its use required a positive action so it is unique in this measure. During the intervention the importance of the probe was a subject of directed attentional focus for the Experimental group and whilst its use was demonstrated during the TAU intervention at the relevant stage in the procedure, the Experimental group were presented with an exercise where failure to use the probe resulted in failure to identify unsupported enamel and caries at the ADJ hidden beneath. The Experimental group were significantly more likely to use the provided dental probe appropriately when assessing each of the models ($P < 0.05$). It is possible that the activity in the training session emphasising its importance created a more memorable experience for the Experimental group leading to them to employ the probe during the transfer test.

8.4.2 Ranking Activity

The Ranking Activity took the same models that participants had just evaluated in isolation and asked them to rank them in order of the best to worst. It was hypothesised that the Experimental group, having received a more structured intervention, would be more in agreement with each other and would rank the models in a similar order. It was also expected that the deeper understanding of the criteria gained from the structured intervention would mean that the Experimental group would suggest a ranking closer to the rational key order provided by a subject matter expert.

To explore the agreement of rankings within each group Kendall's Coefficient of Concordance (Kendall's W) was calculated for both arms of the study. Both groups showed a 'slight' agreement ($0.00 \leq w \leq 0.2$) (Landis and Koch, 1977) with the TAU group having a Kendall's W higher than the Experimental group by a small amount. Only the TAU group showed statistical significance on their level of concordance which suggests greater reliability of results than identified within the Experimental group, however, whilst statistically significant, this is still considered a low agreement. Low agreement from both groups indicates that a greater degree of variability was present in the rankings submitted. It was surprising that the TAU group slightly outperformed the Experimental group as it was hypothesised that the deconstructed tasks would result in a higher level of group agreement but given the difference between the Kendall's W of the two arms was not significant ($P > 0.05$) this could be as a result of the small sample or a lack of sensitivity in the measurement instrument.

The rankings provided by each arm of the study can be further explored by considering the average rankings. Table 8.5 lists the average rankings for each of the models; a lower mean rank indicates that it is more favoured by the participants and is typically ranked higher. The Experimental group demonstrated a preference for the Diamond as the highest ranked model, whereas the TAU group rated the Diamond and Square equally. This suggests that the Experimental group were more able to accurately identify the "best" preparation (discussed bellow) than the TAU group. Collectively, the remaining models were mostly ranked in the order suggested by the rational key (see below) by both groups. However, the TAU group demonstrated a superior standard deviation of mean rankings, suggesting they were in more agreement as to the ranks, reinforcing the results of the Kendall's W calculations.

The range of the mean rankings (difference between the average best rank to the average worst rank) for the Experimental group was 1.7 compared to the range for the TAU group of 1.5 suggesting a greater distinction between the best and the worst ranked models for the Experimental group. However, the stronger preference of the

Experimental group towards the Diamond model is perhaps artificially stretching this range as very little distinction in the average rankings could be identified between models ranked below the Diamond.

It is possible that the lack of discrimination between models across the study could be as a result of there being insufficient differences between them to make the correct rankings discernible to the participants. Equally, it may be as a result of the participant's lack of experience or weaknesses in the intervention itself that contributed to the lack of an ability to discriminate. In this experimental context a balance must be struck between if the differences were more pronounced the level of challenge would be reduced and the likelihood of the participants identifying the correct order increase rendering the rankings as an unreliable discriminator of performance. Furthermore, as a test of transfer it is important that the test is representative of the subtlety present in the target context or the test will present context-irrelevant easiness and compromise its validity (Messick, 1989).

Further insight can be found by considering a pairwise comparison of the individual models across the two arms of the study as shown in Tables 8.6 and 8.7. The Diamond is the most consistently highly ranked model by both the Experimental and TAU groups, in both cases, ranking above every other model at least 10 times. Beyond the Diamond model, there is clearly a lot of uncertainty as to the correct rankings with models being ranked in many different places by both arms of the study. This visually demonstrates the low Kendall's W scores and supports the idea that the level of discrimination a novice could bring to the ranking task was low.

Finally, the marginals matrices shown in Tables 8.8 and 8.9 demonstrate how many times a model was ranked in a given place. The Experimental group strongly agreed that that Diamond was the best model, ranking it 1^{st} 10 times and 2^{nd} twice. The TAU group were more split, ranking the Diamond as the best 8 times and 2^{nd} 4 times. The TAU group were also more varied in their identification of the best model. With the exception of the Triangle (which was the most frequently 2^{nd} placed model) participants in the TAU group suggested every model as the best performance, whereas the Experimental group only suggested the Diamond or Rectangle as the best.

The scores achieved by the participants in reference to the expert-provided rational key showed broadly equivalent performance across the two arms of the study. The TAU group achieved a mean score/difference of 11.38 compared to 12.92 of the experimental group (a lower difference is superior) indicating that the TAU group slightly outperformed the Experimental group. However, the measures here assume that raters have a common understanding of criteria for ranking and there is a clear outlier identifiable in the Quartile-Quartile plot (see Figure 8.4) achieving a score 3x worse than the group average due to having submitted rankings the exact opposite of the rational key. Reviewing the recording of this participant's response, the participant provides a rationale behind their rankings which implies that there was confusion in their understanding of the task. Upon review, it appears that the participant understood "next stage of the procedure" to be in reference to the placement of a restoration, rather than "Managing the caries at the Pulpal Floor". Whilst other participants did not share this interpretation of the task it may be worth considering how such misunderstandings could be mitigated in future work. Taking this strict interpretation of the task, the inverse order is not an unreasonable ranking to suggest. However, this is the answer to a different question, therefore this outlier was excluded from the analysis and the calculations re-performed. After exclusion, the mean score achieved by Experimental group improves to $10.6\dot{6}$, now outperforming the average performance of the TAU group. However, a Two Sample t-test does not indicate a statistically significant ($p > 0.05$) difference between the two groups (either when including or excluding the outlier) in the difference of their rankings with respect to the rational key.

Having now identified an outlier, it is worth returning to the level of group agreement measured using Kendall's W above. Removing the outlier increases the agreement found within the Experimental group to W=0.2652778. This is considered a 'fair' agreement ($0.20 \leq w \leq 0.4$) (Landis and Koch, 1977) and achieved statistical significance (p-value = 0.01265502). If the outlier is accepted as such, this brings the above results of the ranking task into line with existing literature and now demonstrates significance of both the Experimental and TAU arms of the study and suggests that these results support the hypothesis that the deconstructed learning approach promotes a greater shared understanding of the concepts presented as demonstrated via the superior group agreement scores. It should be noted that the small sample size means that the removal of an outlier can disproportionately influence results, however, this participant misunderstood the task so their inclusion equally has a disproportionate influence on results.

Recordings of the participant ranking activities were further explored to investigate the rationale underpinning noteworthy rankings including: 'perfect' scores, participants whose rankings were 'off by one', the participants who ranked the 'worst' model as the 'best' and, those who ranked the 'best' model as the 'worst'.

Participants who achieved a perfect (zero) score on the ranking activity (Experimental n=1, TAU n=2) all demonstrated a sound awareness of the procedure and applied the intended rationale to their rankings. They were all able to correctly reason about the 'worst' preparation, recognising that whilst the ADJ had been successfully cleared, excessive tissue had been irretrievably removed and therefore was an inferior performance to the other preparations which, whilst having deficiencies, could be corrected to achieve a satisfactory result.

Two participants (Experimental n=1, TAU n=1) almost achieved a perfect score, but inverted the order of the 'worst' two models (ranking the Rectangle above the Circle). Both of these participants noted that whilst an excessive volume of tissue had been removed from the Rectangle, the pulpal floor had been prematurely addressed in the Circle. Their reasoning indicated that they prioritised the importance of the correct sequencing of the task in line with the order taught during the intervention (i.e. fully establishing a caries free margin at the ADJ *prior* to addressing the pulpal floor) more highly than the consequent state of the tooth. More widely, performance of correctly ranking the Rectangle in the study was poor and it was the only model to be ranked in all positions by both arms of the study. Whilst five participants recognised this as the worst performance, it is interesting in light of the rationale presented above to recognise that the intended message is not always received. This very much supports the premise that whilst comprehensive educational content depends on a thorough task analysis, a thorough task analysis does not always result in comprehensive educational content (Jonassen et al., 1999, p. vii).

Six participants (including the outlier noted above) ranked the 'worst' model (Rectangle) as the 'best' performance (Experimental n=3, TAU n=3). Unlike the outlier, these five participants did not indicate a misunderstanding of the task, instead they described other features of the preparation including the 'smoothness' of the walls and floor, the lack of unsupported enamel and the clearing of the ADJ to justify their ranking. Some of these mentioned features were not explicitly included in the intervention, however, they were covered during Session 2 of the taught 1st year VR Operative Dentistry module. This could suggest that the prescribed 3 week washout period was insufficient to dissipate the knowledge acquired outside of the intervention.

The ranking activity was arguably the weakest of the measures in this study. The task itself was somewhat abstract and required a level of reasoning that participants were perhaps unprepared for. The interventions did not explicitly focus on being able to compare and evaluate performances therefore the measure is perhaps not fully constructively aligned with the taught material. Future studies should ensure that the measures are manifest in the intervention, however, care must be taken to ensure that bias is not inadvertently introduced: for example, it would be trivial to introduce an exercise that draws attentional focus to comparisons between preparations to assist learners in ranking tasks. However, this could favour the experimental group by "teaching to the test" and produce an unreliable measurement. However, whilst the difference between the groups with respect to the rational key was not statistically significant, after addressing the outlier, there is indication that the experimental group achieved greater agreement which suggests the Experimental approach may yield slightly superior results.

8.4.3 Procedural Activity

The previous tasks focussed on the participant's ability to recall and apply declarative knowledge to a series of models. The procedural task differs by asking participants to verbally describe the steps they would take to complete the goal of *Establishing a Caries Free Margin at the ADJ* from a starting point provided on a new unseen 3D model. This measure focuses on the participant's ability to plan a sequence of actions from retained procedural knowledge to achieve a goal state. This is perhaps the most 'authentic' test of transfer in this study as it requires participants to describe what they would do to complete an operative task. Importantly, this test is detached from any measurement of motor skill: which confounds a measurement of retained knowledge by making it dependent on a participant's ability to execute the intended motions. By using a verbal description, the intentions can be compared under the presumption that the participant would do what they say they would do.

Performance on this task was assessed against a rubric developed from the corresponding stage of the task analysis. Verbs used to discriminate performance were aligned to the descriptors in Bloom's Taxonomy (Bloom and Krathwohl, 1956), with performance being capped at the Analysis level as it would be unrealistic for participants early in the 1st year of their programme to engage in Synthesis or Evaluation of the skill. Scores achieved by participants against this rubric showed a range of [45-70] for the Experimental group and [45-75] for the TAU group. This rubric is valid only to the extent of its intended purpose of discriminating between performances on this specific test (Messick, 1995) and is not intended to provide a globally robust measure of performance for use outside it. By necessity, it is normatively referenced because the *best* a participant could be expected to score is based on the intervention content which is well below the expected standard at a gateway examination. However, as performances were measured across a range of scores in both arms of the study the rubric has demonstrated discrimination so is adequate within the scope of this study.

Performance using this rubric was assessed using the overall mark achieved followed by an exploration of the individual discriminators for any evidence of differences between the arms of the study relating to specific items of procedural knowledge demonstrated by the participants.

The overall score achieved by participants did not show a statistically significant difference between the Experimental group and TAU groups ($p > 0.05$). This presents two possibilities: firstly it could be the case that the rubric was insufficiently sensitive to detect a difference between the performance of the two groups. Alternatively, it could be the case that, regardless of intervention received, participants produced an equivalent performance in the procedural transfer test.

Individual discriminators from the rubric were then examined in turn to explore if there were any differences that might be ascribed to the intervention received, however, no individual aspects of performance produced a statistically significant difference between the performances of either group ($p > 0.05$). Only the "Selection of appropriate burs" discriminator approached statistical significance where the Experimental group received a lower average score (mean = 54.23) than the TAU group (mean = 56.46) indicating the TAU group outperformed the Experimental group. However, with a p-value of 0.06804, this still fell short of the significance threshold. Whilst the TAU achieved a higher mean score, the standard deviation for the TAU group was higher (9.760596) than the Experimental group (5.717719) suggesting more consistent scoring was achieved by the Experimental group. Although, this difference between the two standard deviations was not significant when compared with an F Test ($p > 0.05$).

In the context of this measure, demonstrating no statistically significant difference in performance between the two groups is notable. When measuring the retention of procedural knowledge, the difference between the content of the two intervention sessions is at its greatest. The TAU group received direct instruction on the procedural steps for establishing a caries free margin at the ADJ, being guided through an example case step by step. The number of participants present during the Intervention session was also much smaller than would be in a normal taught session, so it could be argued that the TAU group received the best rendition of current learning and teaching approaches in the VR environment. By contrast, the Experimental group received instruction based on a series of part-task exercises that focussed attention onto individual aspects of performance, but did not re-integrate and demonstrate the performance of the whole activity step-by-step.

That no statistical difference could be detected between the group that received the most 'realistic' training that the simulators can provide and a group who's training deliberately eschewed realism to focus on the underlying knowledge could suggest that the necessity for simulations to be as realistic as possible is questionable and supports this notion as suggested in Towers et al. (2019). These results begin to suggest that to acquire knowledge of the steps in an operative procedure, it is not always necessary for learners to experience the entire procedure as a wholistic experience from start to finish, and instead that knowledge can be acquired with more focussed learning activities. This is not to suggest that leaning should not be reintegrated into performance of the wider task (and it would be likely wise to do so) nor does it preclude the value of a wholistic experience when the intended learning goals require one. However, when deploying novel teaching tools, consideration should be given to which mode of instruction provides the most effective experience and will lead to the maximum transfer to the candidate environment.

8.4.4 Thematic Analysis

At the conclusion of the Transfer Test participants were given an opportunity to share their thoughts on the intervention and their opinions on its effectiveness. Participants in the TAU group were shown images and provided with an explanation of the exercises used in the Experimental arm of the study so that they could share their initial opinions of this alternative approach to learning and teaching.

This section presents a thematic analysis across 4 themes that were identified from participants' comments. Each theme is discussed in detail and related to relevant literature. Where appropriate, discussion is supported by quotations from the participants.

Attentional Focus

Performance is often governed by how well relevant sensory input is detected and acted upon so training that focuses on how performers detect and process sensory information can often lead to improvements in performance (Schmidt and Lee, 2014, p. 64). Many aspects of the intervention were deconstructed and designed to direct the participant's attention towards relevant attributes of the task and these features resonated with the participants and formed a theme around how they contributed to their learning.

The TAU group were not exposed to exercises that explicitly deconstructed the task, however, their intervention session unintentionally offered elements of a deconstructed approach in its delivery. Participants were walked through the goal of establishing a caries free margin at the ADJ step-by-step, keeping the group's progress together. Participants commented how "everyone is focussing on the same aspect of the step at the same time which was good and we also had time to focus specifically on that one aspect" (Participant 24 TAU) and "doing it step by step (...) in a process rather than getting the steps and then doing it all at once (...) was quite helpful" (Participant 13 TAU). This guided approach allowed for those who finished a step more quickly to reflect on their performance and gave time for those working more slowly to immerse themselves in the task and produce a performance they were happy with. This single-step focus means that it could be argued that the TAU intervention had aspects of deconstruction and when coupled with the small group nature of the intervention may have received a more detailed learning experience than would be normal in a larger group setting which may have increased their performance on the transfer tasks.

The Experimental group discussed how the attentional focus features of the exercises contributed to their appreciation of: the motions required (Participant 14, 22 Experimental), awareness of the ADJ (Participants 1, 2, 3, 20, 22 Experimental), being conservative of tooth tissue (Participant 1, 2, 20 Experimental), procedural aspects (Participants 2, 14 Experimental) and the success criteria of the task (Participants 3, 4, 20 Experimental). The comments from the Experimental group covered a much broader spectrum of the underlying concepts than those of the TAU group which focused on the *pace* of delivery rather than the task itself. Perhaps the removal of the extraneous stimuli in the task focussed the Experimental group's attention on the *essence* of the stimulus (Singley and Anderson, 1989, p. 118) and assisted them in identifying what information is irrelevant and what can be disregarded (Hammerton, 1967) which meant that their comments took on a similar character?

There was also broad recognition from the TAU group that the attentional focus features of the Experimental exercises were of value; all reporting positive first impressions based on the overview provided. They appreciated that the Experimental exercises made the appearance of the features clearer and contributed towards understanding (Participants 9, 18, 21, 23 TAU). Participants related this to confusion they encountered in earlier experiences: “when you first start out you’re a bit confused about what you are seeing” (Participant 23 TAU) and that the “arch gets a bit confusing” (Participant 1 Experimental). This may suggest that early in their knowledge acquisition learners appreciate the removal of extraneous cognitive load to support understanding and may not appreciate the full experience that is often provided. Participants appear to recognise the reduced cognitive load that these exercises present and that this simplification reduces the extraneous load (Sweller et al., 1998).

Participant 6 (TAU) commented that enamel was ‘all white’ and they had to figure out how much to remove. The task analysis identified that appropriate removal of enamel was an area of attentional focus and without the guidance provided by the coloured regions of the Experimental activities, it was left to the learner to relate their knowledge to the task itself and put that into practice. As discussed in earlier chapters, some learners may be able to correctly relate the theory and practice but others may not (Resnick, 1975). Being able to perceive *relevant* environmental features is key to being able to understand the nature of the problem (McCloy, 1968) and is often a determinant of how well the skill is executed (Schmidt and Lee, 2014, p. 64). Where a structured and explanatory exercise can assist, there appears to be enthusiasm amongst the participants for that to be offered; Participant 18 (TAU) stated “I think that would be a much more effective way of getting students to understand what the step by step procedure is (...) its definitely a great way to get it into students heads (...) I think it’s much more memorable (...) I would say I’d prefer that a lot”.

Brightly coloured regions intended to draw attention to and demonstrate the relevance of aspects of the task were used throughout the Experimental intervention and attracted much comment. Twelve participants described how they helped highlight the importance of attributes by making them stand out ensuring the learner knew exactly what to look out for and that this helped them understand the content more readily. Participant 18 (TAU) proposed that it is “vital that they [novice students] have it broken down to make it as easy as possible for them because it’s the understanding part that is key”, here recognising the importance of understanding concepts as a separate task to the practice and demonstration of the skill itself. The enthusiasm shown for this feature of the Experimental exercises by so many participants (n=5 Experimental, n=7 TAU) was somewhat surprising. It was expected that the participants would be more critical of bright/unrealistic aesthetics due to the importance placed on realism identified in the literature (Towers et al., 2019). However, this was not the case and only one participant (Participant 24 TAU) expressed a preference for the bright colours to be replaced with a more realistic appearance but even this was disputed by Participant 15 (TAU) who independently suggested (prior to revealing the Experimental approach) that their learning would be supported by “more obvious colours”.

It is worth acknowledging that the participants recruited to this study were very early in their dental degree. This is perhaps evident in the tone of some comments relating to the brightly coloured sections. Any assistance that can assist to identify at the basic level of "what are all these different parts of the tooth" (Participant 11 TAU) are likely to be disproportionately accepted by learners at this stage of the programme. This leads to discussion of the concept of expert blindness because at this early stage a tutor is at their most removed from the level of knowledge held by their learners. Expert blindness recognises that a tutor is not always aware of the difficulties that their learners are facing because the content is so far removed from their skill level and the task so relatively 'basic' that they simply do recognise the extent of support that a novice may require to understand it. Participants in this study illustrate this through a limited notion of the ADJ itself. The task analysis of the previous chapter identified an appreciation of the ADJ as a low-level component in the hierarchy of the cognitive structure of the task and failure to understand these lower-level concepts can lead to the development of an un-sound model of the task (Harlow, 1949). Members of the Experimental group commented that the activities that focussed on an appreciation of the ADJ presented points of view that they had not considered, or had not fully appreciated the significance of (Participants 2 and 22 Experimental). They discussed how "if you had a full tooth you wouldn't be able to see the whole thing" (Participant 20 Experimental) and that the sectional view presented helped them to conceptualise aspects of performance that are not immediately apparent. Conversely, a participant in the TAU arm of the study noted that "there's no real visual representation, we just feel through the dark" (Participant 11 TAU) so perhaps teaching in a way that recognises that even the most basic of concepts can be challenging and takes steps to accommodate them will lead to improved understanding of underlying concepts for junior learners.

Finally, the phrasing "baby steps" was used by a number of participants and that the experimental exercises hold the learners hand. Directing attentional focus towards fewer attributes of the task assists the management of the cognitive load. When a learner's cognitive load is exceeded, learning becomes compromised as there is insufficient cognitive resource to link and consolidate knowledge (Merriënboer and Sweller, 2010) so breaking the task down as done in the Experimental exercises may be a good way to make the content clearer and less overwhelming. If learners are open to this style of exercise and feel that it assists their learning then perhaps exercises such as this could contribute towards ensuring that learners have a comprehensive understanding of the fundamentals before moving on to more wholistic experiences.

Preparing for practice

It is this progression from the fundamentals to subsequent training and clinical practice and how the intervention might support this transition that was identified as the next theme. Whilst the above discussion highlights the value participants placed on illustrating relevant structures with brightly coloured regions, the same feature was discussed more negatively in terms of how it contributes towards preparedness for other contexts. There appears to be a conflict revealed where participants do not initially place a high value on the importance of the simulation's realism but then immediately question this assumption in light of their future progression to clinical cases. For example: "I'm not sure what the realism gives, it doesn't need to be; but you're going to be ultimately doing it, so it needs an element of realism" (Participant 1 Experimental) or "I think maybe at the start not so much. Maybe at the start focus on the different aspects (...) but I think as you get later on I think it is important to see it as realistic just so that you are prepared" (Participant 13 TAU). This perception that the realism will contribute towards preparedness was shared by 6 participants (n=4 TAU, n=2 Experimental) and perhaps shows the importance of face validity in educational exercises. A learner must see the relevance of a learning task in order to see its value (Derous and Born, 2005), so this would suggest that explicitly stating the purpose of such activities to enable learners to appreciate their role in the overall learning journey.

The most prevalent comment relating to the transfer of skills and preparing for practice was how the bright colours used in the Experimental exercises relate to a real-world presentation. Participants reported that they would be "slightly uneasy if I was looking at an actual tooth as opposed to one on the simulator because in the exercise some parts were (...) highlighted in a very obvious colour. So I know that I should work on there because you'd see it and you'd know. But (...) in an actual tooth it won't so you'd need more experience to tell" (Participant 19 Experimental) or more concisely "in real life there is no colour coding" (Participant 6 TAU). These and many other comments suggest a desire to experience an authentic experience, Participant 24 (TAU) went further suggesting that "you can't baby it down too much because you have to be able to transfer those skills onto a real tooth and there can't be such a disparity with what you've practised on". This highlights an interesting point of whether the deconstructed approach is perceived as 'dumbing down' the content? Does this relate to the learners self perception of their maturity as a new university student and a desire to move towards more serious problem-centred andragogy as suggested by Knowles (1973, p.47)?

However, whilst the participants shared these reservations with regard to the bright colours, they also felt that the Experimental exercises were very effective for forming good habits and building confidence. Participants discussed how the deconstructed exercises allowed them to "visualise and understand if I'm doing it right" and that it "offers more reassurance that I'm following the procedures properly" (Participant 11 TAU) and how "it's more about the techniques we are using because if we get into good habits now it will be useful" (Participant 7 TAU). This suggests that participants have a more nuanced perception of the value of realism than the drive towards realistic-as-possible suggested in the literature (Towers et al., 2019). Participants appear to be happy to accept a reduction in the realism if it provides an opportunity for greater learning. But equally, they recognised that these concessions may have consequences for the habits that they form as a result of the experience. For example, Participant 12 (TAU) recognised that it is possible to angle the camera to give views that are not possible in the real world. Participant 5 (TAU) expressed concerns that their difficulties establishing a finger rest in the simulation context might translate to a reduced appreciation of its importance on clinic and Participant 7 (TAU) suggested that familiarity with the feedback from the simulated instruments may need to be unlearned when transferring to the real instruments. These are all valid observations; being able to gain an advantageous view of the tooth being operated upon that is not possible in the candidate context represents a context-irrelevant easiness (Messick, 1995). Learners might opt to take advantage of this functionality to produce a performance that they would not be able to replicate were stricter restrictions imposed and be misled as to their true ability. Furthermore, as noted, the simulated instruments have different characteristics to the real instrument and the fine tuning of motor schemas based on information that is present during training but not replicated in the real world transfer context will lead to an erroneous internal reference of correctness for the task and impaired transfer (Schmidt, 1975; Proteau, 1992). However, the framing of the intervention and the content is of importance here. All participants who raised these issues were part of the TAU arm of the study. Having experienced a simulation that is closer to a real-world experience, any discrepancies between the simulation and the candidate context will be given greater prominence. Conversely, where the goal of the intervention is to educate rather than simulate, realism is given a lower prominence as the feedback from the Experimental group appears to focus more on the content itself rather than the accuracy of the simulation.

This tension between realism and learner-focussed activities lies at the heart of the deconstructed approach. It is possible that as very junior learners, participants were keen to gain exposure to real-world scenarios and this drives a desire to experience realistic simulations. This is confronted by a recognition of their lack of knowledge which suggests that they would benefit from simplified instruction to establish a firm conceptual base to build from. As an approach to education this suggests there may be enthusiasm for both authentic and deconstructed experiences and this is discussed in the next theme.

Preferences for learning and skills development

A recurring theme throughout the comments is that of *understanding* and how the exercises supported and contributed towards the participants' knowledge of the task and its steps. Some comments relate to the use of VR in general and how it is regarded as a means of making content more accessible than is possible using verbal descriptions (Participant 25 TAU) and that it can be used to ensure a shared understanding *of* (Participant 24 TAU) or present new ways of engaging *with* the material that are not possible in the real world (Participant 26 Experimental). However, the main focus of comments related to participants' preferences and thoughts on the respective interventions.

Comments suggested that there is an preference towards the use of the Experimental exercises with 12 participants (n=8 Experimental, n=4 TAU), or almost half, stating that they preferred them. Just 3 participants (n=3 TAU) suggested they would prefer to use wholistic tooth models and the remainder not stating an explicit preference. Participants from both arms of the study commented on how they appreciated the way the Experimental exercises were broken down (Participants 3, 14, 16 Experimental; 12, 23 TAU) as it allowed them to visualise the task (Participants 8, 16, 17 Experimental; 5, 12, 24 TAU) and make more sense of the structure and how it related to the steps (Participants 8, 14, 27 Experimental; 5, 12 TAU). Comments included "I would say that I like it broken down (...) better because it's then aiding you in what to do" (Participant 14 Experimental) and "I think this is better because you can see what happens after each stage and you can see the cross sections so I think it helps us visualise and understand the anatomy of the tooth rather than just doing it on a model" (Participant 8 Experimental).

Participants also felt that the deconstructed approach supported their wider learning, commenting that "it would be a good way of learning alongside our lectures and having both at the same time (..) because you get to see it straight off rather than being taught the theory and then being exposed to this and trying to apply it" (Participant 2 Experimental). The same participant reinforced this point by stating that their participation in the study supported their learning and understanding in aspects of the wider programme.

Participants from the TAU arm differentiated between the two approaches, recognising that "our session felt focussed more on how you're supposed to clear it [the ADJ] and how you deal with a carious lesion rather than actually understanding what it is" (Participant 25 TAU) suggesting that this deconstructed approach helps grasp the underlying concepts.

As discussed above, the enthusiasm for the Experimental approach was somewhat unexpected and suggests an openness to the delivery of learning material that is not simply based around a simulation of the real procedure. Even the 3 participants who preferred the TAU approach stated their preference using less strong terms, using qualifiers whereby they recognised the value in the deconstructed approach but simply preferred the realism that the TAU exercises offered as it represented an exercise closer in appearance to what they will later encounter. This value placed on experiences that bear similarity to future clinical practice was shared even by those who preferred the Experimental approach. Because, even though they preferred the Experimental exercises, this preferences was somewhat transitory. Almost all participants wanted *both* styles of exercise with the Experimental approach being used to establish understanding of the underlying concepts followed by progression to more realistic exercises. Participants felt that the Experimental approach taught the core skills and aided comprehension and then this could be used as a base from which to move on to the wholistic models used in the TAU approach. As suggested by Participant 15 (TAU), "I think that it's pretty good because, as long as that is paired with more traditional way of looking at it, (..) it's quite a confusing thing to understand so seeing it visualised in this way makes it a bit more obvious." and "I think its better for you to understand the basics through this first (...) because at least this way you understand the fundamentals of the different parts of the tooth, the structure, and then you can move on to the whole tooth itself once you've learned everything" (Participant 5 TAU). Comments like this suggest that the participants viewed the deconstructed exercises as scaffolding (Wood et al., 1976) that uses sub-tasks that assist the learner through supportive learning experiences. If these tasks were focussing on the development of motor skills, this approach should be viewed with caution as the sub-task can have different characteristics when carried out in isolation, however, with the focus of the activity on understanding of concepts, the re-integration of the part-task in to the wider task allows it to be *contextualised* by the wider task rather than *modified* by it (Schmidt and Lee, 2014, p. 220).

The Experimental exercises were also valued due to the way they provided a progressive difficulty curve, supporting learners with more simple exercises that build upon each other in complexity. The initial activities of the Experimental approach were regarded as "helpful for putting into perspective the instructions we were given; (...) verbally you explained to us what the unsupported enamel is and where the ADJ is but this would help the student visualise it" (Participant 24 TAU) and that they allowed the participants to "see a lot more detail so I think its a good place to understand what you are doing and when you (...) move on to the models you have an idea what you are doing" (Participant 27 Experimental). These comments illustrate how the exercises introduced the concepts in simple terms and avoided the learner "being thrown straight in to the deep end if you go straight onto models" (Participant 27 Experimental). Participants reported that "it would be helpful to have the first bit, actually understanding it and then moving on to doing it practically, because sometimes you can do stuff practically but not really understand what you are doing. And I feel like the understanding is quite important. I think this approach would help understand it better." (Participant 25 TAU). This suggests that there is a recognition, even at this early stage, that being able to replicate the motions as directed does not necessarily imply the presence of the cognitive knowledge that underpins those motions.

The 'hand-holding' that the Experimental exercises promoted was valued by the participants (Participants 8, 16,17 Experimental; 12 TAU). Participants from the TAU arm of the study felt that the Experimental exercises were "more hand holding you throughout the entire process" and this required less pre-existing knowledge than the TAU exercises because "it is all labelled clearly for you and (...) for intro lessons I think this is easier for people to see it and I know what I'm looking at compared to 'what are all these different parts of the tooth?'" (Participant 11 TAU). This is particularly recognised at an early stage of the course because "you're still learning everything, (...) these procedures are brand new to them and they haven't done anything like this before then. I feel its vital that they have it broken down to make it as easy as possible" (Participant 18 TAU). Participants from the Experimental arm of the study shared similar views: "I think there's a place for this kind of approach, because if you just went straight into doing it in the arch it might be a bit overwhelming because you don't know what you are doing because there's a lot of teeth and you can't see what you are doing properly so to see it, break it down, and just have one tooth: it's like baby steps".

Breaking tasks down as a mechanism to assist understanding agrees with principles of scaffolding (Wood et al., 1976) and cognitive load (Sweller et al., 1998) and participants clearly valued how this made the content more digestible. However, their preferences became moderated in relation to their predictions of future training approaches. As discussed above, Participant 24 (TAU) was concerned about "babying it down" and others cautioned how it may impede skills development, for example, "I think that it would be useful but, as time goes on you need to make a judgement yourself as to if you think you are done because thats like a different skill in itself" (Participant 23 TAU) and "I would prefer not having such a spoon fed approach because that's not going to be followed through the rest of the course (...) I need to be able to develop those skills myself (...) to be able to use the knowledge that I've been given to put that into practice". These comments are not unreasonable, it is well established that the use of training aids can negatively impact performance when those aids are removed (Joseph et al., 2014) so any integration of this approach must be mindful of this phenomena, but equally, it is important to note that the approach is designed to aid with initial understanding of the core concepts, and not as something to be used for repetitive practice towards demonstrating a measure of competence. Once a learner has grasped the underlying cognitive knowledge supported by this style of exercise further 'practice' on these exercises becomes redundant; as Participant 18 (TAU) describes "if you were a 5th year (...) and you've done this procedure lots of times (...), they don't need someone holding their hand the entire way". Under this approach, a clear line must be drawn between the acquisition of knowledge, the practice of the skills and the application of those skills and knowledge to novel cases. It is noteworthy that no evidence was found in the quantitative results above that removing the brightly coloured regions negatively impacted performance in the transfer test used in this study.

However, regardless of the intervention received, all participants reported that they felt that the intervention session prepared them for the transfer tests. Participants from the TAU group felt that the content covered throughout their session included the necessary information to succeed in the transfer test whereas Experimental group participants highlighted Exercise 4 as particularly helpful in their feeling of readiness. It is interesting to note that whilst Exercise 4 was singled out for praise, it was the content of the preceding exercises that gave the participants the knowledge necessary to successfully engage with Exercise 4. Perhaps, this agrees with the value placed on testing as a vehicle to drive learning (Sennhenn-Kirchner et al., 2018) and that value of the preceding exercise is devalued as a result.

It is interesting to compare the fact that all participants felt well-prepared for the transfer tests with their desire to combine the deconstructed exercises with a more realistic version of the task. When framed in concrete terms of their preparedness for a specific test it is possible that participants are able to see the relevance of the knowledge provided by the Experimental exercises and answer in the affirmative. However, when faced with the "unknown" of future tasks they cannot see the same direct applicability and question their preparedness. As 1^{st} year students who are yet to complete their pre-clinical training, the prospect of picking up their dental handpiece and treating their first patient is likely to be a daunting prospect. With high levels of stress being reported throughout the dental degree (Alzahem et al., 2014) any familiarity that can be gained or simulation that can be provided to prepare them for this event is likely to be desirable. Contributing towards preparedness was an early goal for the use of VR simulation and its importance was described in Towers et al. (2022). In this context, regardless of the direct transferability of motor skills simulation still provides value to support the *emotional* preparedness of the learner; and if this were the only benefit of a wholistic simulation experience then it is still of significant value.

VR as a teaching tool: A sub-theme relating to the preferences for learning is the use of VR as a teaching tool and participant opinions of introducing pop-up notifications to aid skills development.

Participants recognised that both VR and Phantom Head modalities had their own strengths and there is a great deal of potential for the future. However, the perception of the value of the VR simulation was somewhat coloured by issues encountered both during the intervention session and during prior experiences of the simulators in timetabled teaching sessions. Eleven participants (n=6 Experimental, n=5 TAU) reported issues with the simulator demonstrating lagginess or behaving 'temperamentally' with a further four participants (n=1 Experimental, n=4 TAU) encountering difficulties with handpiece control.

Issues with system lag result in the on-screen display not keeping pace with the user's motions and intentions. This can result in erratic or unexpected behavior leading to considerable performance degradation (Ware and Balakrishnan, 1994). The exercises developed to facilitate the Experimental arm of the study were known to be complex and to stretch the capabilities of the device used but performance during testing had suggested that it was adequate. However, this does not appear to have been the experience encountered by the participants, with 6 of the 7 participants who described the performance as 'laggy' coming from the Experimental group. The exercises presented to the TAU group were tooth models from the library available on the simulator and participants of this group were more likely to categorise any unanticipated behaviour as 'temperamental' but it is not clear from their comments what the specific issues were.

Regardless of the source, any issues relating to the simulation itself introduce context-irrelevant difficulties (Messick, 1989). These are issues that are present in the training environment which are not relevant to difficulties that would be encountered in the transfer context. These difficulties increase the extraneous cognitive load on the learner and accommodating these deficiencies adds to the total cognitive load, resulting in a reduced capacity for learning (Sweller et al., 1998). It is possible that the skills acquisition of the Experimental group was reduced due to the prevalence of lag-related issues which may have confounded the measurement of the true difference in performance achieved between the two arms of the study. It is important for the use of VR in dental education contexts that system issues are minimised: fundamentally, learners are interested in developing their dental operative skills so issues arising from the simulator or accommodations necessary for its use will be regarded negatively and result in poor perceptions of a device intended to support their skills acquisition.

This study intended to explore the use of pop-up notifications as a mechanism to direct attentional focus to relevant areas of interest and provide context-aware guidance to the learner. This feature was not available in time for the study, so participants were instead asked their opinions on this idea. Most participants who discussed this topic were open to the idea of these notifications. They felt that it might be beneficial for them to work at their own pace, recognising that "everyone worked at different paces and I remember being done and just sat there wondering what to do so maybe if there were pop up boxes" (Participant 4 Experimental) or that it could help make the sessions feel less stressful: "I was like one of the slower people so I feel like if you're a bit behind you panic and feel like you need to go a bit faster and then you do it wrong" (Participant 22 Experimental). Being compelled to rush to keep up will shift their goal towards speed rather than learning which is undesirable (Beilock et al., 2004) so providing individualised feedback relevant to the learner's progress would be a desirable outcome of this approach. Participants were keen on the possibility that the pop-up notifications would provide encouragement on their progress "because there's lots of people in the class, to have something so you know you're doing it the right way, hopefully, yes a bit of encouragement that you're making progress" (Participant 15 TAU) and that it would be "useful because it gives a student a gauge of progress instead of just drilling and not knowing where to stop". Dental operative tasks are of a 'serial' nature with a series of actions linking together, indeed some approach a more 'continuous' nature (Schmidt and Lee, 2014, p9), so judging performance and providing feedback on only the product of the motions means that the feedback may be delivered too late as it is not able to illustrate when a decision was needed or how it contributed to an undesirable outcome. As identified in the task analysis, decisions are made throughout the procedure and directing attention to the relevant attribute of performance at the correct time will result in improved performance (Schmidt and Lee, 2014, p264).

However, whilst participants were open to the idea of the simulator guiding them and providing feedback through pop up notifications, the option of being able to disable them was raised a number of times (Participants 4, 14, 17 Experimental, 23 TAU). As discussed above, participants were keen to test themselves without support because "as time goes on you need to make a judgement yourself as to if you think you are done because thats like a different skill in itself so I think that it would be useful but at the same time you need to know when to leave it alone and when you are done" (Participant 23 TAU).

Finally, participants were keen to defend the role of the tutor in the learning process and expressed reservations at them being replaced by simulator prompts. Some were quite strong in their views, stating that they would "rather have a presentation and talk through it rather than the simulator - I don't think we should rely on technology to teach. I just think people teaching other people is a better connection and you get a better understanding." (Participant 1 Experimental). Other comments suggested that the simulator should provide supplementary feedback, possibly to prompt discussion with the tutor, recognising that the Experimental approach and pop up notifications would not be able to answer any questions or address uncertainties and for that reason the tutor's presence is vital. These comments suggest that participants were open to the idea of the simulators aiding their learning: providing support to aid their progress whilst the tutor is supporting other students. However, they still value the interactions with their tutors so whilst freeing up tutor time was seen as a desirable feature of simulator based feedback (Xia et al., 2013) enthusiasm for this use case is not fully shared by the learners themselves.

Intervention delivery

The intervention was delivered in accordance with the research protocol, however, participants appear to have framed the session as a taught teaching session. This is a positive outcome because participants took the opportunity to provide comment on which elements of the intervention they liked and disliked which can be used to inform future session design that takes advantage of the exercises developed for this work.

The intervention session was structured such that both arms of the study were shown a video recording of a presentation introducing the background to the task, after which the arms of the study diverged. Four participants highlighted the value of the introductory video, reporting that it "was useful because you could compare what you saw on the video to what you saw on the simulator" (Participant 2 Experimental), however, one participant (Participant 15 TAU) found it hard to engage with and struggled at the beginning of the activities. The perceived efficacy of the video content agrees with the conclusions of a systematic review of performed by Gopinath and Nallaswamy (2017) and video content blended with taught sessions is supported as an effective teaching approach by the findings in Iqbal et al. (2022). The comments of the participant who struggled with the video shared views similar to other evaluations of video content in Higher Education where the lack of interactivity leads to disengagement (Boateng et al., 2016) . In the context of this study, the use of a video recording was important to ensure consistency of introductory material across different intervention sessions but in a normal and timetabled teaching session perhaps a more tutor-led session would be better received.

Following the videos participants proceeded to engage with the exercises relevant to their arm of the study. When discussing this, 4 members of the TAU group (6, 7, 9, 23) suggested that they would have appreciated a further demonstration of the procedure to help relate the gestures with the handpiece to those of the tutor (Participants 6, 23), to aid with visualisation (Participant 23) or to just help with understanding the task itself (Participants 7, 9). Both arms of the study received the same introductory video, so it is interesting that only participants from the TAU group requested this additional instruction. Perhaps the structure of the Experimental session with the deconstructed exercises provided the "broken down" step by step interactivity that was needed to understand the procedure whereas the TAU group's walk-through approach did not adequately manage the cognitive load of the task. Being guided through a linear description of a more realistic exercise, it is possible that the distractions of the whole arch (as discussed above) meant the participants were overwhelmed at the volume of information presented leading to reduced germane cognitive load available for learning. Upon reflection, participants may have attributed this to a lack of further (and lower-load) instruction at the outset of the session.

An interesting parallel to this are the comments relating to the simultaneous verbal narration and exercise interaction used during the Experimental session. The intended pop-up notifications would have provided contextual guidance; for example, as the tissue being removed by the learner transitioned from infected to affected the device would display a pop-up message indicating this and drawing attention to relevant features. As this functionality was not available at the time of the study an alternative using verbal narration was provided to describe what the participants should be encountering. Research into working memory suggests that tasks that rely upon different components of working memory should be able to be performed concurrently (Baddeley and Hitch, 1974) and it was believed that the content of the exercise itself would be under the control of the visuospatial sketch pad and the processing of the audio content would be handled by the phonological loop which is responsible for holding acoustic or speech based information (Baddeley, 1992). This split in roles should have allowed for the concurrent processing of these two streams of information. Furthermore, attentional focussing improves performance when a learner's attention is directed towards the product of their motions (Schmidt and Lee, 2014, p264) so it was hoped that providing this verbal information would lead to the desired increase in performance in lieu of the automated feedback. However, three participants from the Experimental arm of the study (Participants 1, 2, 20) reported that they struggled with this approach, suggesting that they found it a challenge to listen to the content and give focus to the activity itself. Others found that their performance on the task was out of step with the information being narrated which meant that there was a mismatch in the auditory content and the activity. This lack of synchronicity was likely to have introduced extraneous cognitive load on the participants and impacted the knowledge acquisition. It is also possible that performing the task was more reliant on informational processing rather than the simple handling of incoming information: if this were the case then it is possible that the two tasks (of understanding the auditory information and carrying out the simulated activity) resulted in interference with each other and performance on both tasks was disrupted (Schmidt and Lee, 2014, p46). This same phenomena may underlie the retrospective request for a demonstration by the TAU group who may have received the walk-through information verbally, but struggled to processes the information concurrently resulting in them feeling that they did not understand the task, seeking additional demonstration to compensate.

Recognising this more broadly in the delivery of a psychomotor skills based programme such as Dentistry is important for educators. Learners may be struggling with processing verbal instructions received concurrently during practice sessions so it is perhaps advisable to ask participants to pause before delivering any contextual information. Had the simulator functionality been available, this pop-up notification could have prompted the learner to pause, read the information and then resume the task having reflected on its relevance.

Finally, four participants complemented the clarity and delivery of their session and 12 participants (n=7 TAU and n=5 Experimental) mentioned how much they enjoyed taking part in the interventions. These unprompted comments, coupled with the fact that the same investigator performed the intervention sessions as administered the transfer test and interviews could suggest some bias may have been present in other comments provided with regard to the study. Whilst every effort was made to separate this fact (including pointing out the importance of their honest opinion, stating that nobody will be offended and that the intervention was delivered following a rigid script) it is still possible that the participants could have felt like they were providing direct feedback to a member of academic staff in relation to their teaching style and the obvious power differential may have contributed to reservations in their responses. However, on balance, participants across the study appear to have provided a range of responses and many of the participants who stated that they enjoyed the sessions also described areas they found challenging or suggested where improvements could be made.

8.4.5 Overall Discussion

Having explored the quantitative and qualitative results of this study individually, the next section will discuss the study as a whole by discussing study limitations and broader topics that span both aspects of the data gathered.

Limitations and Potential Bias

Any overall discussion should begin with a recognition of the limitations of the study and any potential for bias in the methodology. As a single centre study, drawing from a single cohort, using a single manufacturer's simulator, results must be interpreted in line with the confidence that this limited sample can provide. This work is intended to explore the viability of a deconstructed approach to learning and to gauge opinions on exercises in that style. The sample size and confidence measures are scaled appropriately to the addressable population at the host institution and, whilst representative of that population, may not be representative of views or performance at a national or international level. All participants are enrolled at the host institution so based on their pre-application research will share a view that that the environment and teaching style at the University of Sheffield resonated with them. Other institutions will have differing ways of meeting the requirements of the dental degree which will have resonated with a different population who may have returned different results. This is likely to be most acute with the semi-structured interview responses where participants will have framed their responses relative to their experiences to date. For example, a great many participants expressed enjoyment of the session and that they took value from the content, however, this must be viewed in the light of the structure of their programme where operative content is introduced after the timing of the study. Were the programme structured differently or the same study conducted at another institution it is likely that different opinions may have been shared.

Furthermore, participants self-selected for the study which presents the risk of sampling bias. However, almost 1/3 of the available population were recruited to the study and their performance results were normally distributed for most measures which suggests a representative range of abilities were included.

A 3 week washout period was applied between the final timetabled simulation sessions and the interventions, plus a further week before the transfer session. Literature suggests that 3-5 weeks was required to dissipate the effects of the content included in the timetabled course (Wood, 1999), but here it was necessary to retain familiarity with the simulators themselves so the lower bound was selected. However, some participants demonstrated retention of attributes of cavity design that were not covered in the intervention (e.g. attributes of the pulpal floor) and listed these in response to the transfer test questioning. Future studies could apply a number of strategies to avoid this: for example, participants could be directed to only use knowledge acquired during the intervention. This has difficulties because, like a jury being instructed to disregard mis-submitted evidence, the "bell cannot be un-rung" and it may confound the results by introducing uncertainty as to the source of the knowledge and add to the cognitive load to determine if the attribute is permissible. Alternatively, a longer washout period could be used but this would risk participants forgetting how to operate the simulators which could also impede their performance. Neither of these approaches is satisfactory so where possible the experimental sessions should be scheduled prior to any confounding content's timetabled sessions. Unfortunately, this was not possible for this study because the time period between the first simulator familiarisation session and the confounding content of the second session was during an exam period and it was not considered ethical to offer participation during this time. Given the financial incentive offered for completing the study this could have disproportionately incentivised lower income students and adversely impacted their performance in summative assessments.

Measurement should have the maximum dependence on instruction provided and the minimum correlation with pre-instruction knowledge (Ben-Gal et al., 2017). For this study, the overlap between the content of the confounding taught session and the study material was minimal and the measurement criteria focussed on the intervention content. The recall of additional attributes from the taught sessions were not part of the assessment criteria so would not have influenced results. Finally, participants were selected for having completed both sessions of the taught course so all participants had received the same pre-intervention content. Therefore, the risk of this impacting on the results of the study is considered negligible.

The Operative Task and Intervention Exercises

The study recruited from 1st year students on the Bachelor of Dental Surgery degree who, at the time of the study, had just completed their 1st semester of the programme. These participants were selected for their lack of wider knowledge of cavity preparation, however, this choice introduced difficulties whereby the task to explore the deconstructed approach had to require minimal dependance on pre-instruction knowledge (Ben-Gal et al., 2017) to be appropriate for their stage. Establishing a caries free margin at the ADJ has relatively minimal subordinate knowledge and has high face validity as a relevant dental operative task. However, under normal circumstances a learner beginning to prepare cavities would be expected to appreciate what the ADJ is and where it is located: even if lacking practical experience of detecting it operatively. However due to the stage of the programme, these participants required instruction on this subordinate knowledge in order to fully engage with the task of the intervention. However, whilst better familiarity with the anatomy of the tooth would have been desirable for the intervention, as discussed above, these exercises produced insight from the learners and suggest that this style of instruction is valued even for the most simple content.

In the design of the intervention exercises, the focus was to create tasks that shared strategic or conceptual elements with the target context (Schmidt and Lee, 2014, p. 218), optimising for 'transfer appropriate processing' (Salmoni, 1989). This was intended to focus attention on the underlying knowledge that supports later execution by aiding the development of low-level schemas that can be combined to demonstrate understanding (Sweller et al., 1998) and solve problems of that type in a transfer context (Schmidt and Lee, 2014, p. 199). This focus on the underlying knowledge differs from much existing work where learners judge their performance via simulator-provided scores and then seek to progressively improve those scores. These scoring measures are not truly representative of the skill and a high score cannot, with certainty, be assumed to correlate with knowledge of the structure of the task. In this respect, the present study did not seek to maximise the difference in training modalities. The TAU group could have been exposed to an intervention which specified a target area and used this as the basis for instruction. However, a realistic tooth with a carious lesion was used and learners were guided through its removal and, as discussed above, this approach had elements of deconstruction and likely provided a greater cognitive base than if targets were used as the guide to performance.

In future work, there is an opportunity to improve the effectiveness of deconstructed activities. In the present study, the task analysis guided the creation of the exercises, however, it is likely that a greater effect could be measured via a modification to this approach. The Task Description provided a comprehensive guide to the performance of the task. This is an excellent resource to draw upon when developing the marking criteria which could then be used to develop the exercises. Recognising how the task description is manifest in the marking criteria would have drawn out more nuance in the development of the exercises themselves and resulted in an overall approach which is more constructively aligned (Biggs, 2014). When developing the assessment rubric for the measurement of performance, it revealed opportunities for modifications to the exercises that would have likely improved performance. If the assessment embodies the expected (or desired) learning then this should be evident throughout the exercises. In an experimental context, this does raise a question of if the exercises are "teaching to the test" and may bias the outcomes, but as a recommendation for any implementation arising from this work, this sequencing is preferred.

It was intended that the exercises for the Experimental arm would have a "low-poly" aesthetic, that is, have the appearance of a low resolution tooth and clearly not attempt to convince the user that it was intended to be 'realistic'. The intention was that this would mitigate the interference of the Uncanny Valley (Mori, 1970) by virtue of being clearly unrealistic. However, whilst the exercise models were designed with a low resolution design, the simulator applied a smoothing filter which restored a more realistic aesthetic. This could not be disabled so this was accepted and noted for discussion. However, the overall cross-sectional appearance of the exercises appears to have resulted in the desired effect; participants commented on how this made the exercises less realistic but also commenting that it "it doesn't seem like it's too different from what you would actually be doing" (Participant 15 TAU). Given the overall nature of the task, perhaps had the lower resolution aesthetic been possible, the exercises would have had lower face validity and received lower acceptance from the participants.

It was interesting that participants requested both styles of exercises, seeing them as part of a learning journey that values understandability as well as authenticity of the experience. This is particularly noteworthy when considered in relation to the results of the quantitative analysis. Whilst participants wanted both styles of exercise, both arms of the study performed equally well on the transfer tests and where a statistically significant difference was found it favoured the experimental arm. In a VR environment there is an opportunity to separate the learning and the application of knowledge so a series of exercises that are focussed on underlying knowledge could be utilised and then built upon in subsequent exercises towards a more authentic experience. This would provide learners with what they have requested and would be well supported by part-to-whole task practice literature (Schmidt and Lee, 2014, p. 220). However, it is important that these more abstract or analogous tasks are explicitly linked to the problem domain by their tutors so that learners do not have to search for their relevance (Gick and Holyoak, 1980) and alleviate any concerns that the learner would feel unprepared in the target context. Furthermore, in the execution of the more authentic exercise, the tutor should point out the relevance of the experience gained in the part-task exercises so that the knowledge can be more readily transferred (Anderson et al., 1996).

Finally, the unavailable pop-up notifications necessitated adaptations to the intervention. The Experimental exercises were intended to display notifications as participants removed tissue that drew their attention to relevant attributes allowing them to progress through the learning material at their own pace. Likewise, the final Experimental exercise was intended to provide an interactive quiz whereby participants could test their knowledge on the prepared quadrants. As these were not available, the in-line instructional content was moved to verbal narrative supported by slides where necessary and the assessment of the final exercise was changed to an activity whereby participants made their selection and then the correct response discussed with the group. These adaptations meant that the participants did not get the full intended benefit of directed attentional focus during the exercises; effectively both arms of the study were guided through the activities with the main benefit of the Experimental exercises being the reduced cognitive load. Therefore the full benefit of the approach under investigation was not fully realised and the difference in performance resulting from a full implementation may be greater than shown here.

The Transfer Test and Measurement of Learning

This study measures the learning gain of participants via a test of transfer so it is important to state the basis from which this measurement claims its validity (APA, 1954). The series of measurements used in this study satisfy a number of McGrath et al.'s (2015) tests for robust measurements of learning gain: the sample is representative of the cohort (within the stated limitations) and the measure is simulator-agnostic so is comparable across institutions. The test is not longitudinal, which would have increased its robustness, but this was not practical for the present study and a longitudinal measure could have been confounded by further instruction delivered outside of the study. However, the methodology does include multiple measurements which increases its robustness and is scored using a model which is reflective of the domain (Messick, 1995) which meets McGrath's requirement for validity. However, the question should not be if the test is valid, but if the measurements it provides are a sound basis for making inferences (Cronbach and Meehl, 1955) and that the consequences of these measurements do not carry weight beyond what they can bear (Messick, 1989, 1995). For this study the measurements are to compare two alternative interventions and provide insight into future validation of the approach and the measurements are considered acceptable for this use.

The real test of an acquired skill is in the context in which it will be ultimately demonstrated (Salmoni, 1989, p. 218) but clearly it is not ethical or appropriate to ask such novice learners to perform the operative task instructed during the intervention in a real clinical case. So it is more appropriate to test the effectiveness of the simulation training by a measure of how well the participant can generalise and apply the knowledge in an alternative situation (Kozlowski and DeShon, 2020). This is the basis from which the transfer test was developed. The use of 3D printed models creates a repeatable test that can compare performance without being dependent on simulator scores. It separates the performance on the test from the specific model of simulator, allowing for a common test to be used regardless of the device (or lack of device) used for the intervention. Finally, it creates a measure that tests the learner's awareness of the structural and substantive aspects of the task and ensures modality-specific considerations have not been erroneously incorporated into the learner's mental model of the task. This kind of 3D printed transfer test model could form the basis of a future common assessment library containing a standardised collection of deficient preparations with known attributes to be used for assessment of learners' awareness of attributes of performance. The model used in this test were not intended for (and may not be valid for) this purposes, but a series of models that have been appropriately validated by dental professionals could begin to replicate the standardised testing and measurement used to validate aviation simulators (EASA, 2012) and could be used to measure the effectiveness of other simulator based interventions.

Establishing a causal link between an intervention and an improvement is known to be difficult (Colt et al., 2011) and in this study this was further complicated due to the TAU group not being not a *pure* control group. This comparator group were exposed to instruction based on normal teaching methods which are considered effective as part of an accredited programme at the host institution. Therefore, any differences in performance between the Experimental and TAU groups are a measure of the marginal difference between the two interventions. This difference will be smaller and more difficult to detect than would be between a no-treatment control group. However, it is felt that the most meaningful way of measuring the intervention is in comparison to existing approaches and that if a pure control group were used, as is often done elsewhere in the literature, this would result in a biassed result that would not be fully evaluative of the Experimental approach.

To measure performance, the retention and ranking activities were a simple count of declarative facts and the procedural task was a free-form verbal description which was assessed via a rubric. Measuring a difference in learning across a single step of an overall procedure and developing a rubric to do so with such a limited number of differentiators was challenging. This lack of discriminators reduced the ability to differentiate more broadly across the performances. It was a challenge to balance the level of difficulty of the task to be appropriate to the participants but not have to contend with requiring pre-existing knowledge that was still able to provide sufficient intrinsic complexity to differentiate performances at different levels. During the transfer tests, participants appeared to 'settle into the task' and recount a higher level of detail during the procedural activity than they revealed during the retention activity. The process of recounting the steps to describe their intended actions unlocked further knowledge, so it is possible if the order of these tasks were modified that it would result in different performances.

The transfer activities risked introducing the confounder of recognition over recall (Mandler and Rabinowitz, 1981); it is difficult to prompt the participants to provide fuller explanations without inadvertently reminding them of a step or attribute that they wouldn't have volunteered independently but were able to as a result of the prompt or direct question. This was noteworthy where participants linked responses across multiple models, omitting the identification of attributes for a subsequent model because they had only just recounted the same attributes for the previous model. Prompting the participant to include these attributes would have influenced their scores. This was mitigated during analysis by the use of rule based imputation so that participants could be given the benefit of the doubt when a feature was mentioned in one case but omitted in another. However, this did not change any results so was not explored further. Future studies could mitigate this more robustly by asking participants to 'list all of the good and bad attributes', making it clearer that these are discrete focuses that should be fully addressed for each example. Similarly, knowledge structures can be activated by carrying out of the task (Jonassen et al., 1999, p. 194) so perhaps alternative measurements of retention would have been produced if the participant were given the opportunity to carry out the activity using a simulator. This could be achieved by, for example, a think-aloud protocol and describing the steps they were undertaking. Although, this may not be appropriate for novices due to the additional load that thinking-aloud places on the performer (Jonassen et al., 1999, p. 262).

Given that the measurement of learning gain is challenging and that only a marginal difference was available it is perhaps unsurprising that the majority of measurements between the two groups failed to identify a statistically significant difference. However, where one was detected it was invariably in favour of the Experimental arm. Furthermore, the lack of a statistically significant difference is noteworthy as this suggests that the two approaches can be argued to be equivalent. That similar performances can be achieved from participants instructed via an approach not based on a mirror of the realistic situation is at odds with the assertion throughout the literature that the simulation must be as realistic as possible in order to be effective. The results here agree with the work of Biederman and Shiffrar (1987) from a non dental field and suggest that rapid gains in ability can result from engaging with exercises that are based upon *understanding of* rather than *replication of* the task.

It is worth considering, however, if these gains are sufficient to justify the use of the deconstructed approach. The exercises presented to learners cannot be taken out of the context in which they will be used. Tutors have significant experience of delivering training that is close to the real situation and are likely to have been trained using that approach themselves. A realistic simulation allows the tutor to draw upon their real-world experience and relate that to the learner, contrasting the differences between the exercise and reality (Hindmarsh et al., 2014). As discussed above, learners also see significant value in an authentic experience to prepare them for the clinic even though this fear is not replicated in the quantitative results of this study. The experimental exercises are a departure from this, presenting a single-purpose task with reduced face-validity which requires the tutor to teach 'to the exercise', thus reducing the impact of their expertise and requiring them to adopt a less familiar approach. Although, it can also be argued to reduce the dependency on the ability of the tutor, the learning material is structured within the exercise and simply following them should (to the extent of the originating task analysis) ensure that learners have covered all requisite material. Perhaps this would even permit session facilitators to run simulation sessions and free up valuable clinical tutor resource?

8.5 Conclusion

This study presents intriguing early results into the use of deconstructed and cognitive focussed VR exercises in the field of dental education. Across the quantitative measurements the statistically significant results were found to only favour the Experimental group and where no significant difference was detected the performances of the two groups were approximately equal. The Experimental group were instructed using exercises that simplified and explained the concepts but were not exposed to a realistic simulation of the task and yet demonstrated equivalent and often superior performance than the TAU group. This disagrees with the presumption throughout dental simulation literature that to be effective simulation must be as realistic as possible. The qualitative analysis revealed that participants were open to innovative approaches to support their learning and it is not aways desirable for their first exposure to concepts to be in a simulation that closely resembles reality. However, comments show that it is important that these tasks are framed appropriately so that their place in the overall learning journey is signposted and ideally that the journey should culminate in a wholistic experience of the operative task where the knowledge from the preceding activities can be put into practice.

Future work could explore:

- If these same results are replicated using a range of simulators from other manufacturers. The performance of the simulator featured in the qualitative results so it may impact performance, perceptions and the viability of the approach if different hardware were used.
- Given the positive feedback from the participants the pop-up notifications should be implemented and explored to establish if they provide beneficial support in developing understanding of operative tasks.

- The acceptance of presenting the full learning journey: from deconstructed activities, to VR simulated exercises, to phantom head exercises and then exploring if any of these steps are felt to be redundant by the learners.

This is the first instance of VR exercises being employed to specifically teach underlying dental concepts rather than as vehicles for motor-skills acquisition or repetitive practice. It suggests that the use of VR in dental education need not be restricted to simply recreating a digital representation of the phantom head or clinical environment but can also be employed to deliver novel instructional interventions that are not possible or practical using traditional modalities. Results have demonstrated potential to deliver improved learning outcomes for students and that approaches such as this are welcomed by learners to aid their acquisition of knowledge and this suggests that the approach warrants further exploration.

Chapter 9

An Approach for the Development of Deconstructed VR Exercises

The work of the preceding chapters presents the first use of VR as a simulation modality for the deconstruction of a dental operative task using Hierarchical Task Analysis and Task Decomposition. Additionally, they describe the first example of dental VR exercises being used to develop underlying cognitive knowledge instead of providing a wholistic simulation. These novel approaches were trialled in their respective studies and insight from this work is presented below as a series of recommendations for similar work in the future.

9.1 Specification of the educational goal

The first task is to establish the educational goal of the analysis. The below approach assumes that the goal is to develop deconstructed exercises for the use in a VR context. If an analysis is intended to develop teaching material for use in other context, the analyst should be mindful of the limitations discussed previously and consider mitigations such as the inclusion of additional task decomposition step or additional questions to compensate for the differences between the simulator and the real-world context.

- The task being analysed should be considered for its appropriateness. Task analysis is a time consuming process so the intended task should warrant this being analysed in this way by having a significant underlying cognitive component such as judgements to be made, alternative approaches to consider or complex processes to be applied.

- An analysis can also be a useful tool to establish consensus amongst staff members. Different backgrounds, training institutions or even generational differences can all introduce variance in approach which can lead to uncertainty for the learner when they receive seemingly conflicting instruction. Therefore, even if the analysis does not lead to new exercises it can still be a useful activity to achieve consistency.
- The analysis should be carried out as early as possible in the project lifecycle as this is the point where insight from the analysis can have the greatest impact for guiding future efforts. For example, if functionality were being developed to add support for the simulation of a new operative procedure, this would be a good time to apply analysis methods to ensure that the most beneficial functionality is prioritised.
- Finally, staffing beyond just those involved in the analysis should be considered. Developing exercises of this nature requires staff with skills in 3D modelling and it is also worth considering the buy-in of clinical teaching staff. The deconstructed exercises are somewhat a departure from traditional simulated approaches so taking time to introduce and describe how they fit into the wider context is advisable.

9.2 Task Analysis to uncover underlying cognitive knowledge

Having identified a suitable task, the next step is to prepare for and cary out the analysis.

- The first step of preparation is to determine the task decomposition questions. These will guide deeper analysis of the task and ensure that no areas are overlooked. Therefore, any areas of interest should be documented here. For future work with similar goals, the following task decomposition questions are a good starting point for customisation towards a specific project:
  - Please can you describe the overall goal of this step?
  - What equipment did you use during this step and why? Describe the motions used with the instruments.
  - Whilst carrying out this step, where was your attention focussed?
  - What decisions did you need to make and what cues or information did you rely upon when making those decisions?
  - How did you determine that this step was complete?
  - In order to complete this task, what background knowledge did you rely upon?

- What are the identifiable criteria or attributes that a successful execution of this step would demonstrate?
- What are common mistakes novices make during this step?
- What are the consequences of those mistakes?
- Which aspects of this step do learners usually struggle with?
- What are the main differences between carrying out this step on the simulator compared to the corresponding step in the real world?
- Were any of these differences significant and how did you work around them?

- Between 3-5 participants is adequate for the analysis. This number permits the capture of different perspectives and allows any omissions in the descriptions to be cross-compensated but minimises the amount of unnecessary repetition.
- Clearly, the participants should have the task under analysis within their scope of practice but their total experience is less important. An experienced practitioner will describe the task in different terms to one who has more recently qualified. However, both levels of expertise will offer insight into the task and produce a fuller description when combined.
- This extends to participant familiarity with the VR context. Experienced users of VR will be less distracted by the novelty of the simulation and provide a detailed description of their activities. However, it is likely that they are already accommodating some differences and limitations of the modality and may overlook some aspects, whereas someone new to the environment may be surprised by their experiences and provide useful insight.

9.2.1 Conducting task analysis sessions

Having prepared the questions, areas of focus and recruited participants, the task can now be analysed.

- The task analysis session should be conducted one-to-one with the participant and the task analyst. The simulation device used for the analysis should be the same as the one intended for the developed material. This allows the insight of the differences identified during the analysis to be incorporated into future exercise design.
- The simulator should be tested prior to the participant's arrival to ensure that it is fully operational. If this requires the pre-selection of any tools then a completely inappropriate instrument should be selected so that the participant is compelled to make a selection and the rationale for the initial selection is not lost.

- Two cameras should record the session (audio and video): one aimed at the participant hands and another providing a wide-angle of their interactions with the entire device. The wider angle view is preferred if only one camera is available for capturing the session.

- Participants should be provided with background and contextual information about the task and the purpose of the analysis. The contextual information should state any assumptions that have been made for the set up of the task. For example, "it is assumed that an assessment has determined that operative intervention is indicated". As much information as is necessary should be provided, but the participant should be encouraged to mention factors outside of the stage being analysed if it impacts on what they would do.

- To start the task analysis itself, participants should be asked to describe (from memory) what they consider the main 3-6 steps of the task. Each one of these should be written down on a Post-it® note (or similar) and affixed somewhere that is in the participant's view whilst working.

- Next, ask the participant to undertake the actions required to complete the goal described on the first Post-it® note, stopping when complete. The participant should be instructed to think-aloud whilst they are performing the task and include:

  - What they are doing?
  - Why they are doing it?
  - What aspects are they paying attention to that informs their decisions?
  - How do they know when the step is complete?

  The video recording will capture the description itself so the analyst's attention should be on making note of any discrepancies between what is said and what is performed, any areas of uncertainty from the description plus any wider contextual questions such as any pre-existing knowledge that is referenced.

- Between steps, each of the pre-prepared task decomposition questions should be addressed, along with any questions noted by the analyst whilst the participant was working. Some questions will have been covered naturally by the participant during their narrative; it is not necessary to repeat these.

- Once all questions have been addressed the above two steps should be repeated for each of the goals on the Post-it® notes.

- When all steps have been described the analyst should end the session by asking the participant if there is anything else that they should be aware of that was not already included in their descriptions.

9.2.2 Constructing the Task Description

Having conducted the individual task analysis sessions, the Task Description must be created. The descriptions created from individual participants should be verified for accuracy prior to being merged in the creation of the draft Consensus Task Description. This draft consensus description will be presented and discussed with all participants so that a shared and agreed description of the task can be produced.

Constructing the Individual Task Description

- In order to construct the Individual Task Description the video recording of the task analysis session must be reviewed a number of times. Most insight is likely to be found by reviewing the wider angled camera. This can be supplemented with the close-up recording as necessary to more carefully check what was done or if something obscured the view.
- Whilst viewing, a hierarchical graphical model of the goals described for the task can be constructed through the identification of verb phrases or outcome-focussed statements. When encountered, these should be written down on a large sheet of paper. When the next goal is described a decision can be made as to if it is a separate sibling goal or a subordinate goal contributing to one already transcribed. Sibling goals should be written on the same row and sub-goals should be listed below their parent.
- It is a matter of judgement as to what is a sibling goal and what is a sub-goal but the decision can be refined over a number of reviews of the recording. Furthermore, the number of goals is a matter of judgement but, generally, fewer is preferable.
- Once the graphical hierarchy is complete, it is useful to return to the participant's initial high-level goals from the outset of the analysis. Whilst they are likely to be incomplete when compared to their full description, they can be used to confirm if the Task Description is consistent with the participant's initial conceptualisation of the task. If this is not the case then it can be explored with the participant concerned.
- Next, the individual actions that contribute to achieving each goal can be documented in a narrative form in a written document. Task decomposition questions, differences to the real-world task and any supplementary information can be merged into the description or listed under an appropriate heading. The exact structure of this document must be determined by the analyst as is most appropriate for the goals of the analysis.
- Once the Individual Task Description has been transcribed it should be shared with the participant to provide them with an opportunity to comment.

- An optional, but recommended, stage is to arrange a brief follow-up meeting where the description can be presented to the participant. This minimises the risk of them projecting their own understanding onto the analysts work whilst reading. If it is intended that this task analysis will be applied outside of the VR context, this same follow-up meeting could be used to annotate the Task Description with additional detail to accommodate any differences to the real-world task.

The Consensus Task Description

Once all participants have provided feedback on their Individual Task Descriptions the analyst should attempt to unify these descriptions in to a draft Consensus Task Description.

- To do so, all graphical Individual Task Descriptions should be reviewed to identify areas of commonality between descriptions. Where there is an agreed approach, begin to construct an agreed hierarchy by writing these goals onto a large piece of paper. Then review the goals that are in disagreement in more detail in order to judge if they are truly different or simply different ways of describing the same task. If the majority of the participants agree, add the most prevalent description to the draft diagram and note the outlier for discussion during the consensus meeting. If no consensus can be found, select the most appealing description and add it to the draft diagram annotating it with a question mark. Write a summary of all differing approaches for later discussion at the consensus meeting.
- Once a graphical overview of the Consensus Task Description is complete it is time to merge the detailed actions and operations that comprise those goals.
- For each goal, copy the list of tasks from each participant into a new document. Annotate or colour-code the text so that the source of each list can be determined and then attempt to merge the detailed list of tasks to produce a single unified description. It is likely that many tasks will be the same but described using different words. These can simply be merged but again, any unresolvable differences between participants can be noted for discussion later.
- A number of iterations and re-readings of the above steps may be required to produce a satisfactory Task Description that, in the judgement of the analyst, provides cohesive guidance for the task.
- At this point, a draft hierarchical graphical summary of the goals and a document describing the constituent tasks should be available. This draft should now be circulated to all participants and them given sufficient time to review it.
- At a mutually convenient time, all participants should be brought together for a group consensus meeting. The meeting should be recorded to allow for the analyst's

attention to be focussed on the Task Description. During this meeting:

  - All participants should agree to the principles of the Chatham House Rules and acknowledge that they are in effect.
  - The draft document should be verbally presented to the group, pointing out areas of disagreement as they are reached. If all participants confirm that they have already read the document this can be a brief overview.
  - After the presentation, participants should be asked for their initial thoughts and any areas that they wish to discuss or revise.
  - After this discussion, each of the areas of disagreement identified whilst compiling the draft document should be addressed. NB the individuals who disagreed in their descriptions should not be identified; it is for the participants to decide if they wish to reveal themselves as the source of any comments. The analyst can assist with maintaining anonymity by presenting the disagreements as *their* understanding of the descriptions.
  - All discussions should result in an agreed amendment to the draft document as previously shared.

- Following the consensus meeting, all changes should be applied to the final agreed Task Description

- This document can now be shared more widely for further comment if desired.

9.3 Development of exercises

Having described the task including its cognitive underpinnings, a series of exercises that can provide instruction to learners can be developed.

- To do so the Task Description must be thoroughly reviewed to identify the most promising opportunities to create exercises. Areas featuring a significant cognitive component or those where learners are known to struggle are good starting points.

- For each of these areas, ideas for how the concepts can be better explained using VR exercises should be compiled. These proposals should attempt to manage the learner's cognitive load by simplifying concepts or explaining them sequentially. Where possible, the learner's attentional focus should be directed to the aspects of the task that underly expert performance. Sketching proposed exercises and activities to facilitate discussion is encouraged.

- The idea generation stage could be performed by the analyst in isolation but doing so in discussion with subject matter experts and exercise designers will lead to more effective and implementable exercises.

- This list of potential exercises should be reviewed with reference to the limitations and differences between the simulated and real procedure. If the proposed exercise will be limited or confounded by the simulator's functionality (and this cannot be mitigated) then the exercise should be revised or struck off the list.
- Once the list of exercises has been compiled, it is advisable to consider the way that the resultant learning will be measured. For example, if task performance will be assessed then this is an opportunity to develop the assessment criteria. The exercises can then be revisited to evaluate if the taught content will fully prepare the learner for that assessment. This cognitive alignment step may reveal more exercises are required or that modifications might lead to greater impact.
- Finally, the exercises should be explored for any opportunities where activities can be merged into a single exercise to minimise the exercise loading time in the teaching context. Where many separate exercises are unavoidable, the analyst could consider activities that can be brought into the session to keep learners 'on task' whilst they wait.
- Following this iterative approach to exercise design they can finally be created using the appropriate tools as recommended by the vendors of the simulators in use at the host institution.
- After all deconstructed exercises have been created a final wholistic task should be produced to allow the learners to apply and integrate the content that they have been taught in a high face-validity final exercise.

9.4 Measuring learning

As noted above, the criteria used to assess and measure learning can be useful in the design of the exercises. However, these criteria must, to demonstrate validity, be conscious of their intended usage.

9.4.1 In Learning and Teaching

If the measurement is formative and intended for learning and teaching:

- The measurement should be authentic and representative of the knowledge. Proxy measurements that are only demonstrative of the skill in the simulation environment which do not drive transferrable learning should be avoided.
- Any learning measurement taken as a result of exposure to the exercises must be sufficiently robust for the consequences of that measurement. For example,

if the measurement is considered as part of a gateway examination it must be significantly more robust than a formative indicator. The use of transfer tests can contribute to this goal.

- The ultimate aim of learning in a simulation environment is to develop skills that can be demonstrated in other contexts. Transfer tests ask learners to do this so are considered a desirable way of demonstrating that learning has taken place. The transfer test could take many forms but its validity is enhanced if it is portable and can be used in multiple institutions with various facilities.
- The use of 3D printed models for transfer tests is an appealing approach. The test is separated from the simulation context, is reproducible in different locations (without requiring the same simulation hardware), is easily controlled, does not rely upon patients (and confounding variance) and tests the learner's acquisition of the underlying knowledge as detached from their ability to execute the fine finger dexterity skills in a highly face-valid way.

9.4.2 For experimental measurement

Finally, where deconstructed exercises are being evaluated as part of experimental research, in addition to the above the protocol should consider:

- The level of experience of participants in the study. Where a new learning and teaching approach is being explored, it is appealing to recruit learners who are at the same stage of the programme as the targets of the material. These participants are the most similar so can provide the most reliable measure of transfer. However, when a qualitative opinion of the approach is sought, junior learners can lack the wider experience of how dentistry is taught and struggle to compare and contrast to the experimental approach. Therefore, in these cases it could be beneficial to recruit more senior students and forgo the quantitative measure.
- If a quantitative and a qualitative measure is to be taken, it is preferred if these are gathered during separate sessions or after a break so that fatigue does not lead to a failure to ask insightful follow up questions.
- When seeking qualitative feedback from the learners, better responses may be found by conducting a group discussion including the researcher and all participants from each intervention group at once. This could mitigate power differential issues by allowing the larger group of learners to discuss their opinions before presenting their shared, honest, view regarding the intervention. However, this would require participants to attend a further session which may discourage some from taking part in the study.

The above presents a summary of recommendations based on the experiences of the work undertaken in this thesis to guide the implementation of similar work. Further discussion and justification of these recommendations is detailed in the corresponding earlier chapters.

Chapter 10

Discussion

This thesis has presented: a Scoping Review of the uses and applications of VR in dental education; an exploration of wider literature, applying insight from other fields to propose an effective place for VR simulation in pre-clinical education; a novel approach to revealing and deconstructing the underlying cognitive basis of operative tasks; a deconstruction of a caries removal and amalgam cavity preparation task; the creation of novel cognitive-focussed deconstructed exercises; a quantitative and qualitative exploration of their use with a cohort of undergraduate students; and, a series of recommendations for others who wish to use a similar approach in the future. Each of these areas have been extensively discussed in their respective chapters. The discussion that follows will consider the overall limitations of the work, revisit and discuss the study hypotheses and will relate the work carried out to recent literature published during the course of this work.

10.1 Limitations

It must be acknowledged that all work presented in this thesis was carried out at a single institution, using single groups of participants and using a single supplier's simulators. Therefore, any generalisations must consider the impact of these study variables. Whilst recruitment for all studies was sufficient for the stated confidence, recruiting a greater number of participants would have increased the validity of the results presented. Using a single supplier's simulators may also have influenced results and, as discussed in Chapter 2, differences between devices can confound insight being applied elsewhere. However, the premise upon which this work is built is that the simulator should focus on developing underlying cognitive knowledge and this should be demonstrable and transferable to other contexts. Therefore, the use of a transfer test using 3D printed models to evaluate the theory is beneficial as its portability to other institutions contributes towards its validity (McGrath et al., 2015). Different simulators may have revealed different insight into the limitations and suggested different exercises. Task analysis recognises that the skill of the analyst is manifest in the output and whilst a robust approach that controlled as many variables as possible was developed, a different analyst may have made different recommendations.

10.2 Hypotheses

Four hypotheses were presented in Chapter 4 that will be discussed in turn:

Hypothesis H_1 –"A participant who receives instruction based on the cognitive aspects of an operative task using a series of 'deconstructed' part-task exercises on a VR simulator will be able to apply procedural aspects of that knowledge in a transfer task more accurately than one who was taught using a whole-task VR simulation". This was explored by a procedural test in Chapter 8. Across the various measures of this test no statistical difference was found between the two groups, therefore, this hypothesis was not supported by this work. However, if this hypothesis is reframed as "a participant who is instructed via conventional demonstration and practice at the whole-task will perform better than one who only is exposed to separate deconstructed tasks" then surprisingly this, too, is not supported. Significant effort has been devoted to creating better 'simulations' of dental operative tasks, yet the deconstructed activities presented here enabled learners to achieve a level of performance that was not statistically different to those taught conventionally. This may suggest opportunities for VR simulators to offer complementary learning opportunities without being constrained to providing progressively more realistic simulations.

Hypothesis H_2 –"A participant who receives instruction based on the cognitive aspects of an operative task using a series of 'deconstructed' part-task exercises on a VR simulator will have superior retention of the declarative knowledge in a transfer test than one who was taught using a whole-task VR simulation" was investigated by a series of 4 measures of retention for attributes of caries removal and cavity preparation. The experimental group, taught with deconstructed exercises, achieved a higher average score for the total number of correct attributes identified, made the fewest mistakes in identification and performed better on a negatively marked computed measure of retention. The number of mistakes made and the negatively marked measure of retention both achieved statistical significance. Only on the total measure of retention did the treatment as usual group (TAU) perform better, however, this result was not statistically significant and the measure was found to be vulnerable to awarding artificially highly scores if participants guessed answers. Given these results, and within the limitations noted above, this hypothesis is supported.

Hypothesis H_3 –"A deeper understanding of the cognitive aspects of the task acquired from instruction using 'deconstructed' part-task exercises on a VR simulator will allow participants to make better judgements, closer to those of a qualified practitioner, in a transfer task than those instructed using a whole-task VR simulation" was explored by comparing how participants ranked a series of cavity preparations (from best to worst) to a ranking of the same preparations by an expert. After excluding an outlier (who had misunderstood the task) the Experimental group outperformed the TAU group by having a smaller difference between their rankings to those of the expert's but not to a significant degree. Furthermore, the Experimental group were shown to have greater internal agreement within the group. These two factors tentatively support the hypothesis but further evidence is required.

Hypothesis H_4 –"VR exercises based on deconstructed tasks will have a lower face-validity than whole-task and be less well received by novice learners" was explored via a semi-structured interview to gather participant views on the intervention. This hypothesis was very much rejected by the comments of the participants who welcomed the idea of alternative approaches as a mechanism to assist with the understanding of operative concepts. This was somewhat of a surprise given the importance placed on realism in the literature. Participants continue to value a wholistic simulation, and were keen to experience one, but were happy to embrace the experimental exercises into a structured series of preceding steps to prepare them for such an exercise. Based on the responses of participants in this work, this hypothesis is rejected.

10.3 Contemporary Review of the Literature

VR in dental education is a field of active enquiry across the world with the most recent developments informing the results of this work. A new literature search was performed on the 18th December 2024 following the same approach as used for the Scoping Review in Chapter 2 with the additional criteria to exclude publications considered previously. 256 results were returned and all abstracts were manually reviewed with full texts of relevant works retrieved for discussion below.

At the time of searching, no similar works exploring themes relating to deconstructed learning or the use of Task Analysis in the field of VR dental simulation could be found. Relevant works will be discussed in relation to broader themes that guided this work.

10.3.1 Transfer

Recent literature demonstrates a trend for the exploration of the effectiveness of VR training by measuring that learning has occurred using a phantom head. This is a shift from earlier work where the simulator's internal scoring was used to determine if skill acquisition had occured. Some studies presented a single-arm test using a phantom head pre-test followed by a VR intervention then a phantom head post-test (Farag and Hashem, 2022) Others used 2 arms to compare learners trained using a VR intervention to a positive-control group trained using phantom heads, with the experimental measurement being taken using a phantom head (Vincent et al., 2020; San Diego et al., 2022). Others opted for a pure control group (Murbay et al., 2020) or compared performance using the same procedure in VR and phantom head contexts (Hadjichristou et al., 2024). Many of these studies did not take a pre-intervention measurement, but in the instances when this had been done it provided additional insight into the impact of the intervention (Dwisaptarini et al., 2018; Yin et al., 2021).

In many cases, the feedback and assessment used to measure learning focussed on cognitive aspects of performance that related the simulation to real-world criteria for success. For example, Vincent et al. (2020) provided feedback on outline shape, depth, cavity floor regularity and iatrogenic damage; Murbay et al. (2020) provided feedback linked to the SISTA cavity preparation criteria including prepared/overprepared, centredness, contour smoothness etc. and in Vincent et al. (2022) participants received a 10 minute debrief with a tutor. The use of clinically relevant measurements is welcome as it not only provides a measure that can be more readily compared with real-world performance it also provides more authentic feedback that a learner can readily relate to the desired behaviours. However, in all cases the post-tests had confounding influences due to the requirement that the participant demonstrate their learning via motor-performance. A judgement of which participant has 'learned' the most is therefore unreliable as it is only measuring what can be demonstrated. Given the impact of motor skills specificity, there is an element of adaptation being applied to these post-tests so they do not necessarily fully reveal the learner's acquired knowledge from training.

Where the skill is demonstrated in an alternative context, this is showing a 'savings' or 'first-shot' measure (Hammerton, 1967) as a result of the simulation experience. In Vincent et al. (2020) it can be seen that 3 sessions in the VR context produce performance approximately equal to 2 sessions of training using phantom heads. Whist this is not a saving as such, time costs are just one of the considerations: material waste, risk of injury etc. are all relevant measures so whilst the VR training is shown as less time-efficient, that doesn't preclude it being preferable. Other studies demonstrate a 'first-shot' measure, for example San Diego et al. (2022) demonstrates the impact of training across modalities and how students performed when exposed to their first difficult caries removal task.

With the shift in focus to assessing with real instruments using criteria that rely upon underlying knowledge, the VR context has become a training environment to prepare learners for these tests. If this is the case, and learners are being asked to acquire this underling knowledge, is a whole-task simulation needed or, as suggested in this work, would exercises optimised for acquiring this knowledge be preferable?

10.3.2 Simulator Scoring and Feedback

Recent publications show the continued appeal of computed scores based on simulator performance. Published shortly after the Scoping Review, Ria et al. (2018) demonstrated a scoring system for evaluating a learner's cavity preparations. This is based on a measure of the caries removed, healthy tissue removed and an additional penalty for damage to the pulp. Incorporating additional features leads to a more refined score, however, it does not appear to differentiate between the removal of infected and affected tissue or where the caries is located. Whilst this score reveals student progress over time, it is likely to erroneously penalise desirable corrections in cavity design where they occur beyond the designated area. Therefore, the scoring may not promote acquiring the desired behaviours. In more recent work featuring a similar method of assessment it was recommended that new evaluation criteria are necessary to judge performance (Ziane-Casenave et al., 2022).

Attempts to produce more clinically relevant feedback demonstrated positive results from the automated provision of SISTA cavity preparation criteria feedback (Murbay et al., 2020). Participants who received such feedback significantly improved their performance against these criteria. Similar results were presented in (Yin et al., 2021), demonstrating how meaningful automated feedback could drive skills for crown preparation exercises. Work to validate such feedback was demonstrated in (Dixon et al., 2021) where it was shown that the simulator could be used to generate feedback similar to that provided by clinical tutors. These are encouraging findings which suggest that similar positive results could be found in the future using the pop-up notifications that were intended for use in the study presented in Chapter 8.

Assessment and guidance can be provided 'live' whilst the learner is working to provide angulation guides and direct instruction (Vincent et al., 2022). This directly focusses the learner's attention on the factor requiring attention without requiring them to infer the correct action as a result of a retrospective numeric score. Groups taking advantage of this feature were shown to perform better than those without. However, it is important that such guides avoid the learner devolving responsibility for their behaviours to the guide as this is known to reduce performance when it is removed (Schmidt and Lee, 2014, p.275) as was found to be the case in Vincent et al. (2022).

The work presented in this thesis attempted to build upon these ideas of meaningful feedback by framing the dental simulator as, not just a device that could provide corrective feedback upon performance, but one that could be used to teach core concepts. In this way, rather than encouraging the learner to engage in repetitive and iterative practice (which may not transfer to other contexts) instead, focus on knowledge and understanding which can be applied in a modality better suited to developing the corresponding motor performance.

10.3.3 Validity and Validation

Exploring the validity of VR simulators continues to be a focus, with recent publications describing construct validity (Mirghani et al., 2018; Osnes et al., 2021), concurrent validity (Dixon et al., 2021) and both face and content validity (Ba-Hattab et al., 2023). Whilst not stating so explicitly, the predictive validity of measures has also been explored (Urbankova et al., 2022; Al-Saud et al., 2020).

Work to demonstrate construct validity has shown how scores from simulator exercises can capture differences in external ability (Mirghani et al., 2018) such that a greater level of operator experience correlates with an improved score. However, in some cases the criterion measure used awarded qualified clinicians surprisingly low scores. In Osnes et al. (2021) 1st year participants achieved an average of a 20% precision score compared to a qualified clinicians average of 40%. Whilst this is a significant difference between the two groups, showing the measure is able to discriminate, it would be reasonably expected that a measure of such ability would allow a qualified practitioner to score more highly and suggests that the scoring mechanism is not fully capturing the ability of the performer.

Other work has explored how scores from simulator exercises can predict future performance on pre-clinical exams (Urbankova et al., 2022) or more broadly across all exams up to 4th year (Al-Saud et al., 2020). However, this kind of test neglects the wider learning that contributes towards success in dental exams and the influence of external factors such as exam nerves or preparedness. Furthermore, these studies only demonstrate a weak predictive ability, relying upon integration with other (non-simulator based) scores or only show correlations to selected subsequent exams. In similar work, (Joseph et al., 2023) concluded the scoring provided by simulators does not have the ability to distinguish performance reliably.

Furthermore, these scores are largely derived from the performer's ability to demonstrate their skill in a VR simulation yet performance is being compared to assessments taking place elsewhere. The simulator used has an impact upon performance (Hattori et al., 2022) so comparing scores achieved across these different contexts may be unreliable and unrepresentative of skill. Therefore, any resultant consequences of these predictive measures should be proportionate to their robustness (Messick, 1995) and perhaps their use should be restricted, as Al-Saud et al. (2020) suggests, to identifying students who may require additional support.

Importantly, the work presented in this thesis differs from these approaches as it proposes VR simulation as a device to acquire the cognitive knowledge that underpins performance. In this sense, the simulation does not need to be predictive or discriminatory. It is a device on which knowledge is acquired and the effectiveness of the exercises can be evaluated by a measure of if the learner understands the concepts and can apply that knowledge in other contexts.

10.3.4 Perception of VR Simulation

Several studies have explored the perception of and satisfaction with VR simulators in the dental education context. At an institutional level, a recent survey (Serrano et al., 2023) showed that approximately $\frac{2}{3}$ of respondents have integrated VR into their curricula and almost 70% of institutions were satisfied with their devices. However, where studies have explored their acceptability with learners, similar sentiments were shared to those found in Section 8.4.4.

Learners believed that VR simulation was beneficial to their learning (Zafar et al., 2020; Huang et al., 2023; Daud et al., 2023, 2024; Hamama et al., 2024) but that it does not replace the phantom head (Zafar et al., 2020; Daud et al., 2023, 2024) nor does the feedback provide a substitute for tutor feedback (Daud et al., 2024) which agree with the findings in this work. Throughout these studies the most prevalent negative comments related to the realism of the simulation, with responses discussing the feel of the tissue, the sensations of the handpiece and the ability to achieve a finger-rest or orient themselves into the desired position. These comments are somewhat distinct to those found in this work because the realism of the simulation was explicitly not a focus. Whilst participants expressed a desire to experience realistic exercises in addition to what was presented, the experimental exercises were mostly spared such comments because they, with their bright colours and cut-aways, were clearly not attempting to be realistic.

Many of these studies suggest that VR is an alternative to the phantom head so it is perhaps unsurprising that novices, when asked, are unenthusiastic about them being replaced. Phantom head simulation allows a learner to experience using a real handpiece and instrumentation and for a junior student this might be their first opportunity to see themselves in the role of a dental professional. Furthermore, many of these studies recruit pre-clinical students and ask them to evaluate the realism of the VR modality prior to them having experienced the real operative environment. This leads them to compare the VR system to typodonts in a phantom head and suggest the latter is more realistic (Daud et al., 2024). But as discussed earlier, typodonts are themselves unrealistic and VR simulators were designed to replicate the densities of real teeth so providing a comparison to the tactile feel of typodonts is not useful. This is akin to asking someone who has never tasted an orange to evaluate which is more realistic, a Terry's Chocolate Orange® (Terry's Chocolate Co. Ltd., UK) or a bag of orange flavoured Rowntree's Fruit Pastilles® (Nestlé UK Ltd., UK). Both have realistic elements: the Chocolate Orange requires the segments to be separated and is approximately the same size as a real orange. Whereas, a Fruit Pastille is smaller but has appropriate colouring and, lacking the chocolate, tastes more like an orange. To ask someone who has only experienced these confections which is the most realistic may elicit a response (or a preference) one way or the other, but neither is fully representative of the real fruit. Any attempt to make them more alike would be unlikely to produce a desirable combination.

To address this conflict this work positions VR as something different yet complementary to the phantom head which may lead to greater satisfaction with the modality.

10.3.5 Novel uses of VR Simulation

The search for the greatest benefit from VR simulation has led to several novel applications of these devices. These include: combining VR with 3D printed models (Towers et al., 2022), using 3D intra-oral scans to provide learners with an opportunity to practice their clinical case before performing the procedure on their real patient (Serrano et al., 2020; Towers et al., 2022; Hsu et al., 2022) and, taking the opposite approach, providing unrealistic 'caries cubes' to explore the caries removal process (Osnes et al., 2021).

Using 3D intra-oral scans from a real patient to create models that can be applied to the VR context is a powerful learning methodology. This innovation was popular with learners as it allowed them to experiment with alternative approaches to treatment, be fully prepared and led to improvements in patient safety (Serrano et al., 2020; Hsu et al., 2022). This same concept was taken one step further by the incorporation of 3D printed models of the same scan (Towers et al., 2022). Participants here appreciated both modalities, reporting that the VR simulator allowed them to experiment with alternative approaches, then the 3D printed model could be regarded as a 'dry run' of the real procedure where they could focus on other aspects such as where to achieve a finger rest.

'Caries cubes' were presented as an exercise through which caries removal could be taught in a more controllable way. Performance in these exercises was found to discriminate between experts and novices hinting at a diagnostic capability (Osnes et al., 2021). These exercises presented a simplified cubic form from which a novice could extract simulated caries. This reduces the cognitive load of the task by facilitating comprehension of the carious spread without introducing extra information such as the influence of the morphology of the tooth. This differs from the work presented in this thesis as it does not provide instruction on how to approach the caries removal task itself. However, these exercises would make an excellent partner to this work as it could represent an interim step to integrate caries removal knowledge prior to culminating with a fully simulated tooth.

All of these approaches provide the learner with a complementary learning platform that works hand-in-hand with the phantom head environment to promote enhanced skills and knowledge acquisition without unnecessarily competing with it.

10.3.6 Simulation Hardware

The simulator hardware and its functionality continue to be an area of focus, both to validate features of the systems and to explore new developments as they become available.

The value of stereoscopic 3D vision has been explored in three publications, that assess quantitative differences in performance (Kaluschke et al., 2023; Ali et al., 2024) and learner perception (Nassief et al., 2024). There is some disagreement in the importance of stereoscopic vision showing that it improves performance (Ali et al., 2024) or that non-stereoscopic vision is associated with greater improvement (Kaluschke et al., 2023). However, regardless of its impact on performance, more experienced learners preferred it to be available (Nassief et al., 2024).

The work presented in this thesis did not use stereoscopic vision. Its absence was noted by the dental professionals during the task analysis, agreeing with Nassief et al. (2024), but it was not raised during the transfer study. It is possible that the experimental focus on the underlying knowledge meant that this was less of a concern, however, this was also the case for the TAU group. The discrepancy was perhaps because the participants were junior students and, as concurs with the findings from Nassief et al. (2024), this group is less aware of the benefits of stereo vision so were less inclined to mention it.

Wider developments in simulator hardware have also been presented. The availability of high quality consumer-grade 3D headsets and accessible platforms for the creation of VR content appears to have led more groups to explore VR in dental education (Reymus et al., 2020; Mansoory et al., 2022; Rodrigues et al., 2023; Azhari et al., 2024). These developments have explored visualisation (without a haptic interface) of root anatomy (Reymus et al., 2020), the neutral zone and tooth placement (Mansoory et al., 2022) and the development of new desktop simulations (Rodrigues et al., 2023; Azhari et al., 2024).

Using the classifications in Chapter 2 these new devices would be categorised as Desktop PC or Haptic Desktop devices but with an immersive display. However, a more recent commercially available simulator, the SIMtoCARE Dente® (Vreeland, The Netherlands) has appeared in the literature (Ba-Hattab et al., 2023; Serrano et al., 2023; Daud et al., 2023, 2024) that extends the definition of the Dental Skills Trainer.

The SIMtoCARE Dente® provides arguably the closest ergonomic representation of the clinical environment seen in a VR device to date. It provides a physical phantom head which permits teeth to be projected into its mouth when viewing through a transparent screen. This builds upon the devices described earlier by allowing the learner to view their own hand, achieve an accurate finger rest and colocate the position of the instrument and model they are interacting with. The accuracy of the colocation has been shown to have an impact upon performance (Kaluschke et al., 2023) so this is a positive development in simulator design.

However, even despite these enhancements learners still share negative responses when questioned about the realism of these latest devices (Ba-Hattab et al., 2023; Daud et al., 2023, 2024). Criticism focuses on the tactile appreciation of the relative hardness of the tooth structures, depth of perception, omissions from the simulation such as water spray or cheek retraction and more surprisingly the ability to position themselves and manoeuvre around the head to achieve the desired operating position. Perhaps, as perceived by learners, these devices are closer but still not close enough. Perhaps they will never be close enough because they are are 'just a simulation'. To repeat verbatim the observation from Chapter 2: "by striving to create more and more realistic representations of a tooth the slight imperfections and differences from the real thing become the focus of criticism at the expense of recognising the potential of the (perhaps hundreds of) factors that the simulator simulates well".

10.3.7 Summary

The usage of VR in dental education remains mixed. There is a prevalence for their use in pre-clinical education and a focus on their application to the development of motor skills. Their original goals to overcome the difficulties of finding appropriate extracted teeth, the limitations of plastic teeth to replicate realistic experiences, lack of tutor time to provide feedback and the subjectivity of tutor assessment are still as relevant over 10 years after they were listed in Xia et al. (2013). Likewise, the three roles identified in Towers et al. (2019) remain and there is still uncertainty as to if these devices should aim to provide an accurate simulation, a diagnostic or evaluation tool for the identification of struggling students or to be a teaching tool to develop a learner's understanding.

This work proposes that their most effective role is a teaching tool to develop a learner's understanding and in this role they can best contribute to the original goals:

- By placing VR simulation as an *additional* learning environment, the greatest educational outcomes will be found by combining their use with phantom heads.
- In pre-clinical education, the VR context can scaffold knowledge and understanding so that single-use typodont use is minimised and scarce supplies of appropriate extracted teeth are reserved for consolidation and application prior to clinical experience.
- In later learning, the VR continues to scaffold learning but provides a low-consequence environment for experiential learning without promoting repetitive practice to achieve competence.

- By avoiding proxy-measurements of learner performance and instead offering learning environments that promote authentic feedback and instruction to provide learners with the knowledge needed to succeed.
- To work hand-in-hand with the tutor to support the learner in developing their knowledge so that the tutor can add more detailed context and support for taking those skills to a transfer context.

In time, with technical advancements, it may be appropriate to revisit the provision of a wholistic simulation but with present technology the cognitive-focussed, deconstructed, approach mitigates the limitations and criticisms levelled at the wider adoption of VR in dental education and gives an opportunity for them to fulfil their real aim: to help learners.

Chapter 11

Conclusion

This chapter presents conclusions drawn from this work and makes recommendations for dental educators and simulator suppliers:

- Motor skills research (Section 3.2.3) and student opinion (Sections 8.4.4 and 10.3.4) do not support the premise that VR simulation should replace traditional phantom-head simulation.
- The theory of motor skills specificity should be integrated into teaching and its impact should be recognised in educational goals when asking a learner to demonstrate skills in new environments. Consequently, course material should optimise transfer to the target context (Section 3.2).
- The learner experience can be enhanced through a reduction in extrinsic cognitive load by ensuring all involved in teaching share an agreed approach to operative tasks (Section 5.8.4). Task Analysis can be applied to produce a consensus and reveal any divergence to promote discussion leading to an agreed institutional approach.
- Task Analysis is a powerful tool for uncovering the underlying cognitive basis of tasks which can be used to develop new learning materials. It is a time consuming and labour-intensive process so its use should be proportionate to the importance and/or complexity of the educational goal (Chapter 5).
- Deconstructed tasks manage the complexity of learning and facilitate the optimisation of transfer to other contexts. Focussing on the underlying knowledge of a task presents opportunities to explain concepts in new and engaging ways and can mitigate many of the limitations of the VR representation. This work has demonstrated that such exercises can be effective (Section 8.3) and are welcomed by learners (Section 8.4.4).

- A deconstructed approach to learning can result in equivalent performance at a procedural transfer task to traditional teaching approaches (Section 8.3.3). The premise that realism is essential for VR simulation is disputed.
- Results suggest that deconstructed learning can result in superior retention of declarative facts when compared to traditional teaching approaches (Section 8.3.1)
- A learner trained with deconstructed tasks may demonstrate judgement closer to that of a qualified practitioner but further research is required to confirm this (Section 8.3.2).
- Finally, VR in dental education is the amalgamation of many fields of expertise. Simulator vendors, clinicians and educationalists should all collaborate on developments to ensure that the greatest benefit is delivered using these devices (Chapter 2).

11.1 Recommendations for Dental Educators

Based on the work presented in this thesis the following recommendations are made to dental educators seeking to engage with this learning modality in their curricula:

- A learner will demonstrate different levels of competence when performing the same procedure in the VR and phantom head contexts. Performance in one context is not necessarily a predictor of the other (Section 3.2).
- There is a need to ensure that assessment in the VR context is truly representative of the learner's ability. Academic progression should not prevented by context-irrelevant-differences or difficulties of the modality (Section 2.7).
- Learning and teaching approaches used in the phantom head context should not be replicated in the VR context. These modalities are different, each with their own strengths and weaknesses and the re-use of existing approaches is not considered optimal (Sections 2.4 and 10.3.6). It is likely that the best outcomes will be achieved by using both modalities together in a complementary manner.
- The VR context provides the ability to create new and unique learning exercises to explore educational concepts that would be impractical in a phantom head simulation environment. Developing 3D modelling skills to augment the supplier-provided VR learning exercises will allow these opportunities to be explored to their fullest (Chapter 7).
- A learner attempting an operative procedure in a new context will be actively adapting their knowledge and skill. At a motor skills level they are attempting

that procedure for the first time. In supporting these learners, facilitate the transfer by helping them to recall relevant knowledge and relating the experience to the training environment (Section 3.2).

11.2 Recommendations for Simulator Suppliers

Based on the work of this thesis the following recommendations are made to simulator suppliers in the future development of their devices:

- 'Familiarity' exercises such as cutting out crosses and lines should be viewed as such. These exercises establish familiarity with the simulator's operation and a learner's performance on them only serves to demonstrate ability within the context of the simulator itself. Promoting these exercises as a mechanism to develop fine finger dexterity that will contribute to performance elsewhere is not supported by the literature (Section 3.2.3).
- Motor skills specificity should be embraced and used as an impetus to focus efforts on educationally-driven functionality that promotes learning and understanding. This could include exercises or functionality to highlight relevant structures or draw specific attention to noteworthy features. Such enhancements will benefit learners by promoting transfer of skills and knowledge to other contexts and provide an opportunity for VR simulators to make a different and complementary contribution to phantom head learning (Chapter 7).
- A Task Analysis should be performed prior to developing new functionality. This process is necessary to a to identify the features that underly expertise and ensure that appropriate and timely clinical input is provided. This will ensure that the attributes that guide performance are not omitted from the simulation and ensure that these same attributes are not devalued or overlooked by the learner during the corresponding step in the real world procedure (Section 5.1).
- Simulator generated feedback should be meaningful; mapped to the clinical context being simulated and should avoid context-irrelevant scores such as percentage removal statistics. Feedback to the learner should be relatable to the target context: information provided should aid improvements in their performance both in the simulation context and when carrying out the procedure elsewhere (Sections 2.6 and 2.7).

11.3 Future Work

The work of this thesis has revealed numerous opportunities for further research that can build upon and extend the work presented here.

- The Task Analysis approach described in Chapter 5 could be applied to further operative procedures in order to explore their cognitive base and document teaching approaches in use.
- This Task Analysis approach could be adapted to focus on uncovering the deficiencies and differences in a VR simulator's representation of operative tasks. Such a task description could be used to inform future developments and enhancements and guide their most appropriate integration into the curriculum.
- The impact of the simulator used for a Task Analysis could be explored by following a similar approach using alternative systems and comparing the resultant task description with that presented in this thesis.
- The deconstructed exercises explored in Chapter 8 could be further evaluated by presenting them to learners at a different stage in the programme. It would be interesting to explore perceptions of senior students and ask if they agree that such exercises would have assisted their knowledge acquisition.
- The deconstructed exercises could be adapted to be executed on other simulators to explore the impact of a specific supplier's devices upon findings.
- It would be interesting to explore a full realisation of the approach discussed by presenting a cohort of learners with a structured pathway consisting of: Deconstructed VR exercises, progression to wholistic VR exercises and culminating in a Phantom Head exercise. This could explore both learner performance as a result of the pathway and their perceptions of the approach. Such a pathway could also integrate other technologies such as 3D intra-oral scans and 3D printed models to bridge the virtual and real worlds.
- The deconstructed approach should be applied to other operative tasks to explore its effectiveness more broadly.
- The effectiveness of a fully implemented context-aware pop-up notification system should be explored to build on the findings of this work. This could measure the impact of this innovation, especially with regard to self-paced learning and how the removal of tutor guidance is perceived by learners.

Declarative knowledge underpins skill and developing and refining the judgement to correctly apply those skills is a valuable opportunity for VR in dental education. Perfecting motor performance using real handpieces is a skill best delivered with phantom head simulation. However, an equally important role is to develop the understanding of what should be done, where attention should be focused and what to look out for are excellent candidates to take advantage of the features of VR.

> *"A superior pilot uses his superior judgement to avoid situations which require the use of his superior skill"*
>
> – Frank Borman

If VR dental simulation can facilitate the development of this judgement in dental students, then this will lead to better decision making, which will lead to better clinicians and ultimately contribute towards better operative outcomes.

Bibliography

Abernethy, Bruce and Joanne M. Wood (2001), "Do generalized visual training programmes for sport really work? an experimental investigation." *Journal of Sports Sciences*, 19, 203–222.

Adams, Anne E., Wendy A. Rogers, and Arthur D. Fisk (2013), "Skill components of task analysis." *Instructional Science*, 41, 1009–1046.

Adams, Jack A (1971), "A closed-loop theory of motor learning." *Journal of Motor Behavior*, 3, 111–150, URL `https://doi.org/10.1080/00222895.1971.10734898`.

Al-Saud, Loulwa M., Faisal Mushtaq, Richard P. Mann, Isra'A Mirghani, Ahmed Balkhoyor, Richard Harris, Cecilie Osnes, Andrew Keeling, Mark A. Mon-Williams, and Michael Manogue (2020), "Early assessment with a virtual reality haptic simulator predicts performance in clinical practice." *BMJ Simulation and Technology Enhanced Learning*, 6, 274–278.

Al-Saud, Loulwa M., Faisal Mushtaq, Isra'a Mirghani, Ahmed Balkhoyor, Andrew Keeling, Michael Manogue, and Mark A. Mon-Williams (2017), "Drilling into the functional significance of stereopsis: the impact of stereoscopic information on surgical performance." *Ophthalmic and Physiological Optics*, 37, 498–506.

Alessi, Stephen M (1988), "Fidelity in the design of instructional simulations." *Journal of Computer Based Instruction*, 15.

Ali, Huda Al, Sarah Nassief, Ashley Towers, James Field, and Nicolas Martin (2024), "The value of stereoscopic three-dimensional vision on dental students' performance in a virtual reality simulator." *Journal of Dental Education*.

Alzahem, Abdullah M., Henk T. Van der Molen, Arwa H. Alaujan, and Benjamin J. De Boer (2014), "Stress management in dental students: A systematic review." *Advances in Medical Education and Practice*, 5, 167–176.

Anderson, J, ed. (1981), *Cognitive skills and their acquisition*. Lawrence Erlbaum Associates.

Anderson, John R, Lynne M Reder, and Herbert A Simon (1996), "Situated learning and education." *Educational Researcher*, 25, 5–11.

Annett, J, D Duncan, R. B. Stammers, and M. J. Gray (1971), *Task Analysis*. Her Majesty's Stationery Office.

Annett, John (2004), "Hierarchical task analysis." In *The Handbook of Task Analysis for Human-Computer Interaction* (Dan Diaper and Neville A. Stanton, eds.), 67–82, Lawrence Erlbaum Associates.

APA (1954), "Technical recommendations for psychological tests and diagnostic techniques." Technical report, American Psychological Association.

Arksey, Hilary and Lisa O'Malley (2005), "Scoping studies: Towards a methodological framework." *International Journal of Social Research Methodology: Theory and Practice*, 8, 19–32.

Azhari, Amr Ahmed, Yara Ashy, Lolwa Abulhamail, Haitham Sharshar, and Walaa Magdy Ahmed (2024), "Integrating virtual–reality into restorative dentistry training: A study on the acceptability and effectiveness of tooth preparation simulations." *Journal of Dental Education*.

Ba-Hattab, Raidan, Dilek Helvacioglu-Yigit, Lamyia Anweigi, Tayeb Alhadeethi, Mahwish Raja, Sundus Atique, Hanin Daas, Rebecca Glanville, Berkan Celikten, Kaan Orhan, and Kamran Ali (2023), "Impact of virtual reality simulation in endodontics on the learning experiences of undergraduate dental students." *Applied Sciences (Switzerland)*, 13.

Baddeley, Alan (1992), "Working memory." *Science*, 255, 556–559.

Baddeley, Alan D. and Graham Hitch (1974), "Working memory." In *Psychology of Learning and Motivation* (Gordon H. Bower, ed.), volume 8, 47–89, Academic Press.

Banerjee, Avijit and Timothy F. Watson (2015), *Pickard's Guide to Minimally Invasive Operative Dentistry*, 10 edition. OUP Oxford.

Basque, Josianne, Gilbert Paquette, Beatrice Pudelko, and Michel Leonard (2014), "Collaborative knowledge modelling with a graphical knowledge representation tool: A strategy to support the transfer of expertise in organisations." In *Knowledge Cartography : Software Tools and Mapping Techniques* (Alexandra Okada, Simon J Buckingham Shum, and Tony Sherborne, eds.), 2 edition, 491–517, Springer.

Beilock, Sian L, Bennett I Bertenthal, Annette M McCoy, and Thomas H Carr (2004), "Haste does not always make waste: Expertise, direction of attention, and speed versus accuracy in performing sensorimotor skills." *Psychonomic Bulletin and Review*, 11, 373–379.

Ben-Gal, G., E. I. Weiss, N. Gafni, and A. Ziv (2013), "Testing manual dexterity using a virtual reality simulator: Reliability and validity." *European Journal of Dental Education*, 17, 138–142.

Ben-Gal, Gilad, Liori Katorza, Ervin I. Weiss, and Amitai Ziv (2017), "Testing motor learning curves among dental students." *Journal of Dental Education*, 81, 1171–1178.

Ben Gal, Gilad, Ervin I Weiss, Naomi Gafni, and Amitai Ziv (2011), "Preliminary assessment of faculty and student perception of a haptic virtual reality simulator for training dental manual dexterity." *Journal of dental education*, 75, 496–504.

Berger, Joseph, Susan J. Rosenholtz, and Morris Zelditch (1980), "Status organizing processes." *Annual Review of Sociology*, 6, 479–508.

Bernstein, Nicholai A. (1996), "On dexterity and its development." In *Dexterity and Its Development* (Mark L. Latash and Michael T Turvey, eds.), 1 edition, Psychology Press.

Biederman, Irving and Margaret M. Shiffrar (1987), "Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 13, 640–645.

Biggs, John (2014), "Constructive alignment in university teaching." *HERDSA Review of Higher Education*, 1, 5–22, URL www.herdsa.org.au/herdsa-review-higher-education-vol-1/5-22.

Bloom, Benjamin Samuel and David R. Krathwohl (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals, Volume 1*. University of Michigan.

BMJ (2018), "British medical journal - study protocols." URL https://authors.bmj.com/before-you-submit/how-to-write-a-study-protocol/.

Boateng, Richard, Sheena Lovia Boateng, Raphael Baffour Awuah, Eric Ansong, and Augustus Barnnet Anderson (2016), "Videos in learning in higher education: assessing perceptions and attitudes of students at the university of ghana." *Smart Learning Environments*, 3.

Boer, I. R. De, P. R. Wesselink, and J. M. Vervoorn (2013), "The creation of virtual teeth with and without tooth pathology for a virtual learning environment in dental education." *European Journal of Dental Education*, 17, 191–197.

Bogdan, Crenguta M, Alexandru F Dinca, and Dorin M Popovici (2011), "A brief survey of visuo-haptic simulators for dental procedures training." *Proceedings of the 6th International Conference on Virtual Learning, Cluj-Napoca, Romania, Oct*, 28–29.

Bornstein, Robert F. (1996), "Face validity in psychological assessment: Implications for a unified model of validity." *American Psychologist*, 51, 983–984.

Boud, David and David Walker (1998), “Promoting reflection in professional courses: The challenge of context.” *Studies in Higher Education*, 23, 191–206.

Braun, Virginia and Victoria Clarke (2006), “Qualitative research in psychology using thematic analysis in psychology.” *Qualitative Research in Psychology*, 3, 77–101.

CAA (2011), “Civil aviation authority ppl (a) requirements.” URL `https://www.caa.co.uk/general-aviation/pilot-licences/part-fcl-requirements/ppl-spl-bpl-private-pilot-licences/ppl-as-requirements/`. Civil Aviation Authority.

Chambers, David (2012), “Learning curves: What do dental students learn from repeated practice of clinical procedures?” *Journal of Dental Education*, 76, 291–302, URL `http://doi.wiley.com/10.1002/j.0022-0337.2012.76.3.tb05258.x`.

Charness, Neil and Jamie I.D. Campbell (1988), “Acquiring skill at mental calculation in adulthood: A task decomposition.” *Journal of Experimental Psychology: General*, 117, 115–129.

Charters, Elizabeth (2003), “The use of think-aloud methods in qualitative research an introduction to think-aloud methods.” *Brock Education Journal*, 12.

Cheng, Justin, Jaime Teevan, Shamsi T. Iqbal, and Michael S. Bernstein (2015), “Break it down: A comparison of macro- and microtasks.” *Conference on Human Factors in Computing Systems - Proceedings*, 2015-April, 4061–4064.

Cicciù, Marco, Ennio Bramanti, Fabrizio Signorino, Alessandra Cicciù, and Francesco Sortino (2013), “Experimental study on strength evaluation applied for teeth extraction: an in vivo study.” *The open dentistry journal*, 7, 20–6.

Claypool, Mark and Kajal Claypool (2009), “Perspectives, frame rates and resolutions: It’s all in the game.” In *Proceedings of the 4th International Conference on Foundations of Digital Games*, 42–49, ACM.

Claypool, Mark, Kajal Claypool, and Feissal Damaa (2006), “The effects of frame rate and resolution on users playing first person shooter games.” In *Multimedia Computing and Networking(Proceedings of SPIE): Electronic Imaging 2006*, volume 6071, 6071011–60710111, SPIE.

Colt, Henri G., Mohsen Davoudi, Septimiu Murgu, and Nazanin Zamanian Rohani (2011), “Measuring learning gain during a one-day introductory bronchoscopy course.” *Surgical Endoscopy*, 25, 207–216.

Cormier, Jeremy, Denis Pasco, Cedric Syllebranque, and Ronan Querrec (2011), “Virteasy a haptic simulator for dental education.” *Proceedings of the 6th International Conference on Virtual Learning, Icvl 2011*, 61–68.

Corrêa, Cléber Gimenez, Maria Aparecida de Andrade, Moreira Machado, Edith Ranzini, Romero Tori, and Fátima de Lourdes Santos Nunes (2017), "Virtual reality simulator for dental anesthesia training in the inferior alveolar nerve block." *Journal of Applied Oral Science*, 25, 357–366.

Cronbach, Lee J. and Paul E. Meehl (1955), "Construct validity in psychological tests." *Psychological Bulletin*, 52, 281–302.

Daud, Alaa, Manal Matoug-Elwerfelli, Hanin Daas, Daniel Zahra, and Kamran Ali (2023), "Enhancing learning experiences in pre-clinical restorative dentistry: the impact of virtual reality haptic simulators." *BMC Medical Education*, 23.

Daud, Alaa, Manal Matoug-Elwerfelli, Amina Khalid, and Kamran Ali (2024), "The impact of virtual reality haptic simulators in pre-clinical restorative dentistry: a qualitative enquiry into dental students' perceptions." *BMC Oral Health*, 24.

Dawson, Luke J, Kathryn Fox, Mark Jellicoe, Elliot Adderton, Vince Bissell, and Callum C Youngson (2021), "Is the number of procedures completed a valid indicator of final year student competency in operative dentistry?" *British Dental Journal*, 230, 663–670.

de Boer, Ilse R., Maxim D. Lagerweij, Merlijn W. de Vries, Paul R. Wesselink, and Johanna M. Vervoorn (2017), "The effect of force feedback in a virtual learning environment on the performance and satisfaction of dental students." *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, 12, 83–90.

de Boer, Ilse R., P. R. Wesselink, and J. M. Vervoorn (2016), "Student performance and appreciation using 3d vs. 2d vision in a virtual learning environment." *European Journal of Dental Education*, 20, 142–147.

Dees, James W. (1966), "Accuracy of absolute visual distance and size estimation in space as a function of stereopsis and motion parallax." *Journal of Experimental Psychology*, 72, 466–476.

Derous, E. and M. Ph. Born (2005), "Impact of face validity and information about the assessment process on test motivation and performance." *Le travail humain*, 68, 317.

Deshpande, S, S Nahavandi, James Mullins, M B Khan, and D Creighton (2012), "Enhancing audio-haptic enabled novel dental training platform performance." In *Proceedings of the ASME Design Engineering Technical Conference*, volume 2 A and B, 1371–1376.

Diamond, Marian Cleaves (1996), "The brain . . . use it or lose it." *Mindshift Connection*, 1, URL `https://cerebromente.org.br/n11/mente/diamond2.html`.

Diaper, Dan (2004), "Understanding task analysis for human-computer interaction." In *The Handbook of Task Analysis for Human-Computer Interaction* (Dan Diaper and Neville A Stanton, eds.), 5–49, Lawrence Erlbaum Associates.

Dinh, Huong Q., Neff Walker, Larry F. Hodges, Chang Song, and Akira Kobayashi (1999), "Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments." *Proceedings - Virtual Reality Annual International Symposium*, 222–228.

Dixon, J., A. Towers, N. Martin, and J. Field (2021), "Re-defining the virtual reality dental simulator: Demonstrating concurrent validity of clinically relevant assessment and feedback." *European Journal of Dental Education*, 25.

Drury, C. G. (1983), "Task analysis methods in industry." *Applied Ergonomics*, 14, 19–28.

Dwisaptarini, A. P., S. Suebnukarn, P. Rhienmora, P. Haddawy, and S. Koontongkaew (2018), "Effectiveness of the multilayered caries model and visuo-tactile virtual reality simulator for minimally invasive caries removal: A randomized controlled trial." *Operative Dentistry*, 43, E110–E118.

EASA (2011), "Acceptable means of compliance and guidance material to part-fcl 1." European Aviation Safety Agency.

EASA (2012), "Certification specifications for aeroplane flight simulation training devices." European Aviation Safety Agency.

Ericsson, K. Anders and Herbert A. Simon (1980), "Verbal reports as data." *Psychological Review*, 87, 215–251.

Esambe, E, C Mosito, and S Pather (2016), "First-year students' essay writing practices: Formative feedback and interim literacies." *Reading & Writing-journal of the Reading Association of South Africa*, 7.

Esambe, E E (2018), "Teaching research writing at exit-level undergraduate programmes in south africa." *South African Journal of Higher Education*, 32, 382–400.

Escobar-Castillejos, David, Julieta Noguez, Luis Neri, Alejandra Magana, and Bedrich Benes (2016), "A review of simulators with haptic devices for medical training." *Journal of Medical Systems*, 40, 1–22.

EUR-Lex (2017), "Regulation (eu) 2017/852 of the european parliament and of the council." URL `https://eur-lex.europa.eu/eli/reg/2017/852/oj/eng`.

Evans, C., C. Kandiko Howson, and A. Forsythe (2018), "Making sense of learning gain in higher education." *Higher Education Pedagogies*, 3, 1–45, URL `https://doi.org/10.1080/23752696.2018.1508360`.

Eve, Elizabeth, Samuel Koo, Abdulmonem Alshihri, Jeremy Cormier, Maria Kozhenikov, R Bruce Donoff, and Nadeem Karimbux (2014), "Performance of dental students versus prosthodontics residents on a 3d immersive haptic simulator." *Journal of Dental Education*, 78, 630–637.

FAA (1992), “Advisory circular 120-45a.” Technical report, Federal Aviation Administration.

Farag, Abeer and Danya Hashem (2022), “Impact of the haptic virtual reality simulator on dental students’ psychomotor skills in preclinical operative dentistry.” *Clinics and Practice*, 12, 17–26.

Fernández, C, G Esteban, M Á Conde, and F Garcia (2016), “Improving motivation in a haptic teaching/learning framework.” *International Journal of Engineering Education*, 32, 553–562.

Field, J., S. Stone, C. Orsini, A. Hussain, S. Vital, A. Crothers, and D. Walmsley (2018), “Curriculum content and assessment of pre-clinical dental skills: A survey of undergraduate dental education in europe.” *European Journal of Dental Education*, 22, 122–127.

Field, J. C., J. G. Cowpe, and A. D. Walmsley (2017), “The graduating european dentist: A new undergraduate curriculum framework.” *European Journal of Dental Education*, 21, 2–10.

Field, James (2016), *Pre-Clinical Dental Skills at a Glance*, 1 edition. Wiley Blackwell.

Field, James, Jonathan Dixon, Ashley Towers, Rachel Green, Hathal Albagami, Guy Lambourn, Julie Mallinson, Wietske Fokkinga, Jorge Tricio-Pesce, Tin Crnić, and Sibylle Vital (2020), “Defining dental operative skills curricula: An adee consensus paper.” *European Journal of Dental Education*, 405–414.

Fitts, Paul Morris and Michael I. Posner (1967), *Human Performance*, 1st edition. Brooks/Cole Pub. Co.

Fleishman, Edwin A. (1975), “Toward a taxonomy of human performance.” *American Psychologist*, 30, 1127–1149.

Frederiksen, John R. and Barbara Y. White (1989), “An approach to training based upon principled task decomposition.” *Acta Psychologica*, 71, 89–146.

French, Ashley, Leah K. Taylor, and Melissa R. Lemke (2019), “Task analysis.” In *Applied Human Factors in Medical Device Design* (Mary Beth Privitera, ed.), 63–81, Elsevier Inc.

Fugill, M. (2013), “Defining the purpose of phantom head.” *European Journal of Dental Education*, 17, 1–4.

Furnham, Adrian (2000), “The brainstorming myth.” *Business Strategy Review*, 11, 21–28.

Gagne, Robert M. (1962), “The acquisition of knowledge.” *Psychological Review*, 69, 355–365.

Gagné, Robert Mills and Karen L Medsker (1996), *The conditions of learning: Training applications*, 5 edition. Harcourt Brace College Pub.

GDC (2015), "Preparing for practice: Dental team learning outcomes for registration (2015 revised edition)." General Dental Council.

Geng, Jason (2013), "Three-dimensional display technologies." *Advances in Optics and Photonics*, 5, 456–535.

Germans, Desmond M., Hans J W Spoelder, Luc Renambot, Henri E. Bal, Sander van Daatselaar, and Paul van der Stelt (2008), "Measuring in virtual reality: A case study in dentistry." *IEEE Transactions on Instrumentation and Measurement*, 57, 1177–1184.

Gick, Mary L. and Keith J. Holyoak (1980), "Analogical problem solving." *Cognitive Psychology*, 12, 306–355.

Gilbert, Stephen B (2016), "Perceived realism of virtual environments depends on authenticity." *Presence*, 25, 322–324.

Gopinath, Vidhya and Deepak Nallaswamy (2017), "A systematic review on the most effective method teaching dentistry to dental students compared to video based learning." *American Journal of Educational Research*, 5, 63–68.

Gottlieb, Riki, Sharon K. Lanning, John C. Gunsolley, and Judith A. Buchanan (2011), "Faculty impressions of dental students' performance with and without virtual reality simulation." *Journal of Dental Education*, 75, 1443–1451.

Guadagnoli, M, W Holcomb, and M Davis (2002), "The efficacy of video feedback for learning the golf swing." *Journal of sports sciences*, 20, 615–622.

Guion, Robert M. (1980), "On trinitarian doctrines of validity." *Professional Psychology: Research and Practice*, 11, 385–398.

Guthrie, Edwin Ray (1952), *The Psychology of Learning.* Peter Smith Pub Inc.

Hadjichristou, Christina, Dimitrios Symeonidis, and Evangelia Dimitriadou (2024), "Haptics vs typodonts for crown preparation in undergraduate dental student education." *European Journal of Prosthodontics and Restorative Dentistry*, 32, 175–182.

Hake, R R (2007), "Should we measure change? yes - evaluation of teaching and student learning in higher education." *Evaluation of Teaching and Student Learning in Higher Education*, 1–57.

Hake, Richard R. (1998), "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses." *American Journal of Physics*, 66, 64–74.

Hamama, Hamdi, Ka Yan Harrison, and Sukhdeep Murbay (2024), “Benefits of using virtual reality in cariology teaching.” *BMC medical education*, 24, 1051.

Hammerton, M. (1967), “Simulators for training.” *Electronics and Power*, 13, 8.

Harlow, B Y Harry F (1949), “The formation of learning sets.” *Psychological Review*, 56, 51–65.

Harmon-Jones, E. (2012), “Cognitive dissonance theory.” *Encyclopedia of Human Behavior: Second Edition*, 543–549.

Hattori, Akitaka, Ken ichi Tonami, Jun Tsuruta, Masayuki Hideshima, Yasuyuki Kimura, Hiroshi Nitta, and Kouji Araki (2022), “Effect of the haptic 3d virtual reality dental training simulator on assessment of tooth preparation.” *Journal of Dental Sciences*, 17, 514–520.

Heiland, M, a Petersik, B Pflesser, U Tiede, Rainer Schmelzle, K.-H Höhne, and H Handels (2004), “Realistic haptic interaction for computer simulation of dental surgery.” *International Congress Series*, 1268, 1226–1229.

Henry, F. M. (1968), “Specificity vs. generality in learning motor skill.” In *Classical Studies on Physical Activity* (R. C. Brown and G. S. Kenyon, eds.), 331–340, Prentice Hall.

Hindmarsh, Jon, Lewis Hyland, and Avijit Banerjee (2014), “Work to make simulation work: 'realism', instructional correction and the body in training.” *Discourse Studies*, 16, 247–269.

Holding, Denis H., ed. (1989), *Human skills*, 2nd ed. edition. Wiley.

Hsu, Min Hsun, Hui Wen Yang, Chia Min Liu, Chun Ju Chen, and Yu Chao Chang (2022), “Clinical relevant haptic simulation learning and training in tooth preparation.” *Journal of Dental Sciences*, 17, 1454–1457.

Huang, Yue, Yingwen Hu, Unman Chan, Pengyu Lai, Yueting Sun, Jun Dai, Xin Cheng, and Xuesong Yang (2023), “Student perceptions toward virtual reality training in dental implant education.” *PeerJ*, 11.

Ioannou, Ioanna, Edmund Kazmierczak, and Linda Stern (2015), “Comparison of oral surgery task performance in a virtual reality surgical simulator and an animal model using objective measures.” In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 5114–5117, IEEE, URL `http://ieeexplore.ieee.org/document/7319542/`.

Iqbal, Azhar, Kiran Kumar Ganji, Osama Khattak, Deepti Shrivastava, Kumar Chandan Srivastava, Bilal Arjumand, Thani Alsharari, Ali Mosfer A. Alqahtani, May Othman Hamza, and Ahmed Abu El Gasim Abdelrahmandafaalla (2022), “Enhancement of skill competencies in operative dentistry using procedure-specific educational videos (e-learning tools) post-covid-19 era—a randomized controlled trial." *International Journal of Environmental Research and Public Health*, 19.

Jamieson, Earle S., James H. Chandler, Peter R. Culmer, Michael Manogue, Mark Mon-Williams, and Richard M. Wilkie (2015), "Can virtual reality trainers improve the compliance discrimination abilities of trainee surgeons?" *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2015-Novem, 466–469.

Jonassen, David H., Martin Tessmer, and Wallace H. Hannum (1999), *Task Analysis Methods for Instructional Design.* Routledge.

Joseph, David, Jeanne Davril, Éric Mortier, Jean Marc Martrette, Nguyen Tran, Pascale Corne, and Marin Vincent (2023), "Distinguishing skill levels with haptic simulation in restorative dentistry: Myth or reality?" *European Journal of Dental Education.*

Joseph, David, Jean Philippe Jehl, Pablo Maureira, Cyril Perrenot, Neal Miller, Pierre Bravetti, Pascal Ambrosini, and Nguyen Tran (2014), "Relative contribution of haptic technology to assessment and training in implantology." *BioMed Research International*, 2014.

Kaluschke, Maximilian, Myat Su Yin, Peter Haddawy, Siriwan Suebnukarn, and Gabriel Zachmann (2023), "The effect of 3d stereopsis and hand-tool alignment on learning effectiveness and skill transfer of a vr-based simulator for dental training." *PLoS ONE*, 18.

Kalyuga, Slava, Remy Rikers, and Fred Paas (2012), "Educational implications of expertise reversal effects in learning and performance of complex cognitive and sensorimotor skills." *Educational Psychology Review*, 24, 313–337.

Kelley, T. L. (1927), *Interpretation of educational measurements.* Macmillan.

Kikuchi, Hirono, Masaomi Ikeda, and Koji Araki (2013), "Evaluation of a virtual reality simulation system for porcelain fused to metal crown preparation at tokyo medical and dental university." *Journal of dental education*, 77, 782–92, URL http://www.ncbi.nlm.nih.gov/pubmed/23740915.

Kim, Laehyun and Se Hyung Park (2006), "Haptic interaction and volume modeling techniques for realistic dental simulation." *Visual Computer*, 22, 90–98.

Kirwan, B. and L. K. Ainsworth, eds. (1992), *A Guide to Task Analysis.* Taylor & Francis.

Knowles, Malcolm (1973), *The Adult Learner: A Neglected Species.* Gulf Publishing Company.

Kolesnikov, Maxim, Miloš Žefran, Arnold D. Steinberg, and Philip G. Bashook (2009), "Periosim: Haptic virtual reality simulator for sensorimotor skill acquisition in dentistry." *Proceedings - IEEE International Conference on Robotics and Automation*, 689–694.

Konukseven, E Ilhan, M Ercument Onder, Erkan Mumcuoglu, and Reha Sukru Kisnisci (2010), "Development of a visio-haptic integrated dental training simulation system." *Journal of dental education*, 74, 880–891.

Kozhevnikov, Maria, David W. Schloerb, Olesya Blazhenkova, Samuel Koo, Nadeem Karimbux, R. Bruce Donoff, and Jairo Salcedo (2013), "Egocentric versus allocentric spatial ability in dentistry and haptic virtual reality training." *Applied Cognitive Psychology*, 27, 373–383.

Kozlowski, Steve W. J. and Richard P. DeShon (2020), "A psychological fidelity approach to simulation-based training: Theory, research and principles." *Scaled Worlds: Development, Validation and Applications*, 75–99.

Kuh, George D. (2003), "What we're learning about student engagement from nsse: Benchmarks for effective educational practices." *Change: The Magazine of Higher Learning*, 35, 24–32.

Kusumoto, Naoki, Taiji Sohmura, Shinichi Yamada, Kazumichi Wakabayashi, Takashi Nakamura, and Hirofumi Yatani (2006), "Application of virtual reality force feedback haptic device for oral implant surgery." *Clinical Oral Implants Research*, 17, 708–713.

Landis, J R and G G Koch (1977), "The measurement of observer agreement for categorical data." *Biometrics*, 33, 159–174.

Larson, Elaine L., Sameer J. Patel, David Evans, and Lisa Saiman (2013), "Feedback as a strategy to change behaviour: The devil is in the details." *Journal of Evaluation in Clinical Practice*, 19, 230–234.

Lee, Cliff, Sheetal R. Asher, Supattriya Chutinan, German O. Gallucci, and Hiroe Ohyama (2017), "The relationship between dental students' assessment ability and preclinical and academic performance in operative dentistry." *Journal of Dental Education*, 81, 310–317.

Lee, Timothy D. and Richard A. Magill (1983), "The locus of contextual interference in motor-skill acquisition." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9, 730–746.

Li, Yi, Sanyuan Zhang, and Xiuzi Ye (2017), "Penalty-based haptic rendering technique on medicinal healthy dental detection." *Multimedia Tools and Applications*, 76, 10825–10835, URL `http://link.springer.com/10.1007/s11042-016-3985-y`.

Lian-Yi, Chen and Abe Toshiyuki (2003), "A dental training system using virtual reality." *Computational Intelligence*, 430–434.

Lin, Xiaodong, Cindy Hmelo, Charles K. Kinzer, and Teresa J. Secules (1999), "Designing technology to support reflection." *Educational Technology Research and Development*, 47, 43–62.

Lockwood, Sally N, Sharon M Compton, Jacqueline L Green, and Kari Rasmussen (2018), "Dental hygiene students' clinical skill acquisition: Activity theory and the use of videos." *Journal of dental hygiene : JDH*, 92, 15–21.

Luchins, Abraham S (1942), "Mechanization in problem solving: The effect of einstellung." *Psychological Monographs*, 54.

Luciano, Cristian, Pat Banerjee, Thomas DeFanti, and Sanjay Mehrotra (2004), "Realistic cross-platform haptic applications using freely-available libraries." *Proceedings - 12th International Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, HAPTICS*, 282–289.

Luciano, Cristian, Pat Banerjee, and T. DeFenti (2009), "Haptics-based virtual reality periodontal training simulator." *Virtual Reality*, 13, 69–85.

Makinson, O F and W R Hume (1982), "Pen and instrument grips and clinical performance in dental students." *Australian Dental Journal*, 27, 86–90.

Mandler, George and Jan C Rabinowitz (1981), "Appearance and reality: Does a recognition test really improve subsequent recall and recognition?" *Journal of Experimental Psychology: Human Learning and Memory*, 7, 79–90.

Mannan, Sam (2005), "Human factors and human error." In *Lees' Loss Prevention in the Process Industries* (Sam Mannan, ed.), third edition, 14/1–14/92, Butterworth-Heinemann.

Mansoory, Meysam Siyah, Seyyed Mohsen Azizi, Fakhrosadat Mirhosseini, Danial Yousefi, and Hedaiat Moradpoor (2022), "A study to investigate the effectiveness of the application of virtual reality technology in dental education." *BMC Medical Education*, 22.

Mansueto, Michael A., Hatem A. Abdulkarim, Walid R. Thabet, and Stephan J. Haney (2010), "The chamfer finish line: Preclinical student performance using different bur designs." *Journal of Dental Education*, 74, 612–617.

McCloud, Scott (1994), *Understanding Comics: The Invisible Art.* HarperCollins.

McCloy, C H (1968), "A preliminary study on factors in motor educability." *Classical Studies on Physical Activity*, 307–318.

McGrath, Cecile Hoareau, Benoit Guerin, Emma Harte, Michael Frearson, and Catriona Manville (2015), "Learning gain in higher education." Technical report, RAND Corporation (on behalf of HEFCE).

McMahan, Ryan P., Alexander Joel D. Alon, Shaimaa Lazem, Robert J. Beaton, David Machaj, Michael Schaefer, Mara G. Silva, Anamary Leal, Robert Hagan, and Doug A. Bowman (2010), "Evaluating natural interaction techniques in video games." *3DUI 2010 - IEEE Symposium on 3D User Interfaces 2010, Proceedings*, 11–14.

McMahan, Ryan P, Chengyuan Lai, and Swaroop K Pal (2016), "Interaction fidelity: The uncanny valley of virtual reality interactions." In *Virtual, Augmented and Mixed Reality* (Stephanie Lackey and Randall Shumaker, eds.), 59–70, Springer International Publishing.

Medellín-Castillo, H. I., E. H. Govea-Valladares, C. N. Pérez-Guerrero, J. Gil-Valladares, Theodore Lim, and James M. Ritchie (2016), "The evaluation of a novel haptic-enabled virtual reality approach for computer-aided cephalometry." *Computer Methods and Programs in Biomedicine*, 130, 46–53, URL http://dx.doi.org/10.1016/j.cmpb.2016.03.014.

Meltzer, D. E. (2002), "Normalized learning gain: a key measure of student learning. addendum to: The relationship between mathematics preparation and conceptual learning gains in physics: A possible "hidden variable" in diagnostic pretest scores." *American Journal of Physics*, 70, 1259.

Merriënboer, Jeroen J G Van and John Sweller (2010), "Cognitive load theory in health professional education: Design principles and strategies." *Medical Education*, 44, 85–93.

Messick, Samuel (1989), "Meaning and values in test validation: The science and ethics of assessment." *Educational Researcher*, 18, 5–11.

Messick, Samuel (1995), "Validation of psychological assessment." *American Psychologist*, 50, 741–749.

Microsoft (2024), "Microsoft excel for mac." Microsoft Corporation.

Militello, Laura G. and Robert J.B. Hutton (1998), "Applied cognitive task analysis (acta): A practitioner's toolkit for understanding cognitive task demands." *Ergonomics*, 41, 1618–1641.

Miller, George A. (1994), "The magical number seven, plus or minus two: Some limits on our capacity for processing information." *Psychological Review*, 101, 343–352.

Miller, Robert B (1953), "A method for man-machine task analysis." Technical report, The American Institute for Research.

Mirghani, I., F. Mushtaq, M. J. Allsop, L. M. Al-Saud, N. Tickhill, C. Potter, A. Keeling, M. A. Mon-Williams, and M. Manogue (2018), "Capturing differences in dental training using a virtual reality simulator." *European Journal of Dental Education*, 22, 67–71.

Mislevy, R. J., L. S. Steinberg, F. J. Breyer, R. G. Almond, and L. Johnson (1999), "Cognitive task analysis with implications for designing simulation-based performance assessment." *Computers in Human Behavior*, 15, 335–374.

Mohamed, N. F F and D. A. Luke (2002), "Virtual reality on the www: Simulating class i procedures in restorative dentistry." *2002 Student Conference on Research and Development: Globalizing Research and Development in Electrical and Electronics Engineering, SCOReD 2002 - Proceedings*, 360–363.

Mori, Masahiro (1970), "The uncanny valley." *Energy*, 7, 33–35.

Murbay, Sukhdeep, Prasanna Neelakantan, Jeffrey Wen Wei Chang, and Shadow Yeung (2020), "Evaluation of the introduction of a dental virtual simulator on the performance of undergraduate dental students in the pre-clinical operative dentistry course." *European Journal of Dental Education*, 24, 5–16.

Mylon, Peter, Roger Lewis, Matt J Carré, and Nicolas Martin (2016), "Evaluation of the effect of medical gloves on dexterity and tactile sensibility using simulated clinical practice tests." *International Journal of Industrial Ergonomics*, 53, 115–123.

Nassief, Sarah, Huda Al Ali, Ashley Towers, James Field, and Nicolas Martin (2024), "Dental students' perceptions of the use of two-dimensional and three-dimensional vision in dental education using a three-dimensional haptic simulator: A qualitative study." *Journal of Dental Education*.

O'Keefe, M, V Wade, S McAllister, I Stupans, and T Burgess (2016), "Improving management of student clinical placements: insights from activity theory." *BMC MEDICAL EDUCATION*, 16.

O'Keefe, M, V Wade, S McAllister, I Stupans, J Miller, T Burgess, A LeCouteur, and L Starr (2014), "Rethinking attitudes to student clinical supervision and patient care: a change management success story." *BMC MEDICAL EDUCATION*, 14.

Olson, Gary M, James D Herbsleb, and Henry H Rueter (1994), "Characterizing the sequential structure of interactive behaviors through statistical and grammatical techniques." *Human-Computer Interraction*, 9, 427–472.

Onnasch, L., J. Bürglen, S. Tristram, and H. Kolrep (2016), "Break it down! a practitioners' software tool for human factors task analysis and human error identification." *Proceedings of the 32rd Conference of the European Association for Aviation Psychology – Thinking High and Low: Cognition and Decision Making in Aviation*.

Osborn, J W and J W Summitt (1998), "Extension for prevention: is it relevant today?" *American Journal of Dentistry*, 11, 189–196.

Osnes, Cecilie, Alice Duke, Jianhua Wu, Paul Franklin, Faisal Mushtaq, and Andrew Keeling (2021), "Investigating the construct validity of a haptic virtual caries simulation for dental education." *BMJ Simulation and Technology Enhanced Learning*, 7, 81–85.

Osnes, Cecilie and Andrew Keeling (2017), "Developing haptic caries simulation for dental education." *Journal of Surgical Simulation*, 4, 29–34, URL http://www.journalsurgicalsimulation.com/contents/volume4/JSS170006/.

Owen, Harry (2016), *Simulation in Dentistry and Dental Hygiene*, 431–450. Springer International Publishing, URL https://doi.org/10.1007/978-3-319-26577-3_11.

Paas, Fred G W C and Jeroen J G Van Merrienboer (1994), "Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach." *Journal of Educational Psychology*, 86, 122–133.

Paas, Fred G.W.C. and Jeroen J.G. Van Merriënboer (1994), "Instructional control of cognitive load in the training of complex cognitive tasks." *Educational Psychology Review*, 6, 351–371.

Papavassiliou, Giorgos and Gregoris Mentzas (2003), "Knowledge modelling in weakly-structured business processes." *Journal of Knowledge Management*, 7, 18–33.

Paquette, Gilbert, Michel Léonard, Karin Lundgren-cayrol, Stefan Mihaila, Denis Gareau, and Source Journal (2006), "Learning design based on graphical knowledge-modelling." *Journal of Educational Technology & Society*, 9, 97–112.

Perani, D., F. Fazio, N. A. Borghese, M. Tettamanti, S. Ferrari, J. Decety, and M. C. Gilardi (2001), "Different brain correlates for watching real and virtual hand actions." *NeuroImage*, 14, 749–758.

Pickering, James D. (2017), "Measuring learning gain: Comparing anatomy drawing screencasts and paper-based resources." *Anatomical Sciences Education*, 10, 307–316.

Piso, E. (1981), "Task analysis for process-control tasks: The method of annett et al. applied." *Journal of Occupational Psychology*, 54, 247–254.

Plessas, Anastasios (2017), "Computerized virtual reality simulation in preclinical dentistry: Can a computerized simulator replace the conventional phantom heads and human instruction?" *Simulation in Healthcare: Journal of the Society for Simulation in Healthcare*, 12, 332–338.

Polanyi, Michael (1958), *Personal knowledge : towards a post-critical philosophy*. Routledge.

Proteau, Luc (1992), *On the specificity of learning and the role of visual information for movement control.*, 67–103. North-Holland.

Păvăloiu, I B, S A Sandu, S Grigorescu, and G Dragoi (2016a), "Inexpensive dentistry training using virtual reality tools state of the art." In *Proceedings of INTED2016 Conference*, 279–285.

Păvăloiu, Ionel-Bujorel, Radu Ioaniţescu, George Drăgoi, Sorin Grigorescu, and Simona Andreea Sandu (2016b), "Virtual reality for education and training in dentistry." In *The 12th International Scientific Conference eLearning and Software for Education Bucharest*, 115–120.

R Core Team (2023), "R: A language and environment for statistical computing." URL `https://www.R-project.org/`. R Foundation for Statistical Computing.

Reedy, Gabriel B (2015), "Using cognitive load theory to inform simulation design and practice." *Clinical Simulation in Nursing*, 11, 355–360, URL `http://dx.doi.org/10.1016/j.ecns.2015.05.004`.

Resnick, Lauren B. (1975), "Task analysis in instructional design: Some cases from mathematics." Technical report, Pittsburgh Univ., Pa. Learning Research and Development Center.

Reymus, M., A. Liebermann, and C. Diegritz (2020), "Virtual reality: an effective tool for teaching root canal anatomy to undergraduate dental students – a preliminary study." *International Endodontic Journal*, 53, 1581–1587.

Rhienmora, P., P. Haddawy, P. Khanal, S. Suebnukarn, and M. N. Dailey (2010), "A virtual reality simulator for teaching and evaluating dental procedures." *Methods of Information in Medicine*, 49, 396–405.

Rhienmora, Phattanapon and Kugamoorthy Gajananan (2010), "Haptic augmented reality dental trainer with automatic performance assessment." *Proceedings of the 15th international conference on Intelligent user interfaces*, 425–426.

Rhienmora, Phattanapon, Peter Haddawy, Siriwan Suebnukarn, and Matthew N. Dailey (2009), "A vr environment for assessing dental surgical expertise." *Frontiers in Artificial Intelligence and Applications*, 200, 746–748.

Rhienmora, Phattanapon, Peter Haddawy, Siriwan Suebnukarn, and Matthew N. Dailey (2011), "Intelligent dental training simulator with objective skill assessment and feedback." *Artificial Intelligence in Medicine*, 52, 115–121, URL `http://dx.doi.org/10.1016/j.artmed.2011.04.003`.

Ria, Sama, Margaret J. Cox, Barry F. Quinn, Jonathan P. San Diego, Ali Bakir, and Mark J. Woolford (2018), "A scoring system for assessing learning progression of dental students' clinical skills using haptic virtual workstations." *Journal of Dental Education*, 82, 277–285.

Rodrigues, Pedro, Francisco Nicolau, Martim Norte, Ezequiel Zorzal, João Botelho, Vanessa Machado, Luís Proença, Ricardo Alves, Carlos Zagalo, Daniel Simões Lopes, and José João Mendes (2023), "Preclinical dental students self-assessment of an improved operative dentistry virtual reality simulator with haptic feedback." *Scientific Reports*, 13.

Rosen, Jacob, Massimiliano Solazzo, Blake Hannaford, and Mika Sinanan (2002), "Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model." *Computer Aided Surgery*, 7, 49–61.

Roy, Elby, Mahmoud M. Bakr, and Roy George (2017), "The need for virtual reality simulators in dental education: A review." *Saudi Dental Journal*, 29, 41–47, URL `http://dx.doi.org/10.1016/j.sdentj.2017.02.001`.

Ruthenbeck, Gs and Kj Reynolds (2015), "Virtual reality for medical training: the state-of-the-art." *Journal of Simulation*, 914, 16–26.

Rutledge, Chrystal, Catharine M. Walsh, Nathan Swinger, Marc Auerbach, Danny Castro, Maya Dewan, Mona Khattab, Alyssa Rake, Ilana Harwayne-Gidansky, Tia T. Raymond, Tensing Maa, and Todd P. Chang (2018), "Gamification in action: Theoretical and practical considerations for medical educators." *Academic Medicine*, 93, 1014–1020.

Rystedt, Hans and Björn Sjöblom (2012), "Realism, authenticity, and learning in healthcare simulations: Rules of relevance and irrelevance as interactive achievements." *Instructional Science*, 40, 785–798.

Sabalic, Maja and Jason D. Schoener (2017), "Virtual reality-based technologies in dental medicine: Knowledge, attitudes and practice among students and practitioners." *Technology, Knowledge and Learning*, 22, 199–207.

Salas, Eduardo, Clint A Bowers, and Lori Rhodenizer (1998), "It is not how much you have but how you use it: Toward a rational use of simulation to support aviation training." *The International Journal of Aviation Psychology*, 8, 197–208.

Salmon, Paul, Daniel Jenkins, Neville Stanton, and Guy Walker (2010), "Hierarchical task analysis vs. cognitive work analysis: Comparison of theory, methodology and contribution to system design." *Theoretical Issues in Ergonomics Science*, 11, 504–531.

Salmoni, Alan W (1989), "Motor skill learning." In *Human Skills* (Denis H. Holding, ed.), second edition, 197–227, John Wiley & Sons Ltd.

Salmoni, Alan W, Richard A Schmidt, and Charles B Walter (1984), "Knowledge of results and motor learning: A review and critical reappraisal." *Psychological Bulletin*, 95, 355–386.

San Diego, Jonathan P., Margaret J. Cox, Barry F A Quinn, Jonathan Tim Newton, Avijit Banerjee, and Mark Woolford (2012), "Researching haptics in higher education: The complexity of developing haptics virtual learning systems and evaluating its impact on students' learning." *Computers and Education*, 59, 156–166, URL `http://dx.doi.org/10.1016/j.compedu.2011.11.009`.

San Diego, Jonathan P., Tim J. Newton, Anika K. Sagoo, Tracy Ann Aston, Avijit Banerjee, Barry F.A. Quinn, and Margaret J. Cox (2022), "Learning clinical skills using haptic vs. phantom head dental chair simulators in removal of artificial caries: Cluster-randomized trials with two cohorts' cavity preparation." *Dentistry Journal*, 10.

Saunders, Jeffrey A. and David C. Knill (2003), "Humans use continuous visual feedback from the hand to control fast reaching movements." *Experimental Brain Research*, 152, 341–352.

Schmidt, Richard A. (1975), "A schema theory of discrete motor skill learning." *Psychological Review*, 82, 225–260.

Schmidt, Richard A. and Timothy D. Lee (2014), *Motor Learning and Performance : From Principles to Application*, 5th edition. Human Kinetics.

Segura, Cristina, Diego Halabi, and Nancy Navarro (2018), "Design and validation of a basic dental psychomotor skills test for novice dental students." *Journal of Dental Education*, 82, 1098–1104.

Sennhenn-Kirchner, S., Y. Goerlich, B. Kirchner, M. Notbohm, S. Schiekirka, A. Simmenroth, and T. Raupach (2018), "The effect of repeated testing vs repeated practice on skills learning in undergraduate dental education." *European Journal of Dental Education*, 22, e42–e47.

Serrano, Carlos M., Dirk R. Bakker, Masie Zamani, Ilse R. de Boer, Pepijn Koopman, Paul R. Wesselink, Erwin Berkhout, and Johanna M. Vervoorn (2023), "Virtual reality and haptics in dental education: Implementation progress and lessons learned after a decade." *European Journal of Dental Education*, 27, 833–840.

Serrano, Carlos M., Paul R. Wesselink, and Johanna M. Vervoorn (2020), "First experiences with patient-centered training in virtual reality." *Journal of Dental Education*, 607–614.

Shahriari-Rad, Arash, Margaret Cox, and Mark Woolford (2017), "Clinical skills acquisition: Rethinking assessment using a virtual haptic simulator." *Technology, Knowledge and Learning*, 22, 185–197.

Shea, John B. and Susan T. Zimny (1983), "Context effects in memory and learning movement information." *Advances in Psychology*, 12, 345–366.

Singley, Mark K and John R Anderson (1989), *The Transfer of Cognitive Skill.* Harvard University Press.

Skinner, B F (1961), *The Science of Learning and the Art of Teaching.*, 145–157. Appleton-Century-Crofts.

Snoddy, G. S. (1926), "Learning and stability: A psychophysiological analysis of a case of motor learning with clinical applications." *Journal of Applied Psychology*, 10, 1–36.

Sofronia, Raluca Elena, Arjana Davidescu, and George Gustav Savii (2012), "Development of an efficient haptic system for tooth processing." *Advanced Materials Research*, 463-464, 651–654, URL `http://www.scientific.net/AMR.463-464.651`.

Somrang, Nontapol and Nopporn Chotikakamthorn (2006), "Interactive haptic simulation of dental plaque removal." *2006 International Symposium on Communications and Information Technologies, ISCIT*, 319–322.

SPIRIT (2013), "Spirit (standard protocol items: Recommendations for interventional trials)." URL `https://www.spirit-statement.org`.

Steinberg, Arnold D, Philip G Bashook, James Drummond, Seema Ashrafi, and Milos Zefran (2007), "Assessment of faculty perception of content validity of periosim, a haptic-3d virtual reality dental training simulator." *Journal of dental education*, 71, 1574–1582.

Suebnukarn, S., R. Hataidechadusadee, N. Suwannasri, N. Suprasert, P. Rhienmora, and P. Haddawy (2011), "Access cavity preparation training using haptic virtual reality and microcomputed tomography tooth models." *International Endodontic Journal*, 44, 983–989.

Suebnukarn, Siriwan, Monthalee Chaisombat, Thanapohn S Kongpunwijit, and Phattanapon S Rhienmora (2014), "Construct validity and expert benchmarking of the haptic virtual reality dental simulator." *Journal of Dental Education*, 78, 1442–1450.

Suebnukarn, Siriwan, Piyawadee Chanakarn, Sirada Phisutphatthana, Kanchala Pongpatarat, Udom Wongwaithongdee, and Chanekrid Oupadissakoon (2015), "Understanding information synthesis in oral surgery for the design of systems for clinical information technology." *British Journal of Oral and Maxillofacial Surgery*, 53, 968–975, URL `http://dx.doi.org/10.1016/j.bjoms.2015.08.001`.

Suebnukarn, Siriwan, Peter Haddawy, Phattanapon Rhienmora, and Kugamoorthy Gajananan (2010a), "Haptic virtual reality for skill acquisition in endodontics." *Journal of Endodontics*, 36, 53–55, URL `http://dx.doi.org/10.1016/j.joen.2009.09.020`.

Suebnukarn, Siriwan, Peter Haddawy, Phattanapon Rhienmora, Pannapa Jittimanee, and Piyanuch Viratket (2010b), "Augmented kinematic feedback from haptic virtual reality for dental skill acquisition." *Journal of dental education*, 74, 1357–1366.

Suebnukarn, Siriwan, Nattharat Phatthanasathiankul, Sunantha Sombatweroje, Phattanapon Rhienmora, and Peter Haddawy (2009), "Process and outcome measures of expert/novice performance on a haptic virtual reality system." *Journal of Dentistry*, 37, 658–665.

Suebnukarn, Siriwan, Pawornwan Rittipakorn, Budsara Thongyoi, Kwanwong Boonpitak, Mansuang Wongsapai, and Panu Pakdeesan (2013), "Usability assessment of an electronic health record in a comprehensive dental clinic." *SpringerPlus*, 2, 1–8.

Suskie, Linda (2004), "Re: pre- post testing in assessment." URL `https://lsv.uky.edu/scripts/wa.exe?A2=ASSESS;6e504167.0408&S=`. ASSESS - Assessment in Higher Education LISTSERV.

Swapp, David, Vijay Pawar, and Céline Loscos (2006), "Interaction with co-located haptic feedback in virtual reality." *Virtual Reality*, 10, 24–30.

Sweller, John (1988), "Cognitive load during problem solving: Effects on learning." *Cognitive Science*, 12, 257–285.

Sweller, John, Jeroen J.G. Van Merrienboer, and Fred G.W.C. Paas (1998), "Cognitive architecture and instructional design." *Educational Psychology Review*, 10, 251–296.

Syllebranque, Cédric and Christian Duriez (2010), "Six degree-of freedom haptic rendering for dental implantology simulation." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5958 LNCS, 139–149.

Teras, M (2016), "Inter-professional working and learning: instructional actions and boundary crossing or boundary making in oral healthcare." *Journal of Education and Work*, 29, 614–636.

Thorndike, E. L. and R. S. Woodworth (1901), "The influence of improvement in one mental function upon the efficiency of other functions. (i)." *Psychological Review*, 8, 247–261.

Thyvalikakath, Thankam P., Michael P. Dziabiak, Raymond Johnson, Miguel Humberto Torres-Urquidy, Amit Acharya, Jonathan Yabes, and Titus K. Schleyer (2014), "Advancing cognitive engineering methods to support user interface design for electronic health records." *International Journal of Medical Informatics*, 83, 292–302, URL `http://dx.doi.org/10.1016/j.ijmedinf.2014.01.007`.

Tjiam, Irene M., Barbara M.A. Schout, Ad J.M. Hendrikx, Albert J.J.M. Scherpbier, J. Alfred Witjes, and Jeroen J.G. Van Merriënboer (2012), "Designing simulator-based training: An approach integrating cognitive task analysis and four-component instructional design." *Medical Teacher*, 34.

Toosi, Amirhosein, Mohammadreza Arbabtafti, and Barry Richardson (2014), "Virtual reality haptic simulation of root canal therapy." *Applied Mechanics and Materials*, 666, 388–392, URL `http://www.scientific.net/AMM.666.388`.

Torczyner, Harry (1977), *Magritte: Ideas and Images*. H. N. Abrams.

Towers, Ashley, Jonathan Dixon, James Field, Rachel Martin, and Nicolas Martin (2022), "Combining virtual reality and 3d-printed models to simulate patient-specific dental operative procedures—a study exploring student perceptions." *European Journal of Dental Education*, 26, 393–403.

Towers, Ashley, James Field, Christopher Stokes, Stephen Maddock, and Nicolas Martin (2019), "A scoping review of the use and application of virtual reality in pre-clinical dental education." *British Dental Journal*, 226, 358–366, URL `https://doi.org/10.1038/s41415-019-0041-0`.

Tse, Brian, William Harwin, Alastair Barrow, Barry Quinn, Jonathan San Diego, and Margaret Cox (2010), "Design and development of a haptic dental training system - haptel." *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6192 LNCS, 101–108.

Umemura, Hiroyuki (2014), "Effects of haptic information on the perception of dynamic 3-d movement." *PLoS ONE*, 9, 1–8.

Urbankova, Alice, Miroslav Eber, and Steven P Engebretson (2013), "A complex haptic exercise to predict preclinical operative dentistry performance: a retrospective study." *Journal of dental education*, 77, 1443–50.

Urbankova, Alice and Steven P Engebretson (2011), "The use of haptics to predict preclinic operative dentistry performance and perceptual ability." *Journal of dental education*, 75, 1548–57, URL `http://www.ncbi.nlm.nih.gov/pubmed/22184593`.

Urbankova, Alice, Leena Palomo, and Steven P. Engebretson (2022), "A complex haptic exercise to predict pre-clinic operative dentistry performance: A prospective study." *Journal of Dental Education*, 86, 1628–1633.

USDoE (2003), "Identifying and implementing educational practices supported by rigorous evidence: A user friendly guide." URL `http://www.ed.gov/print/rschstat/research/pubs/rigorousevid/guide.html`. US Department Of Education, Coalition for Evidence-Based Policy.

Vermunt, Jan D., Sonia Ilie, and Anna Vignoles (2018), "Building the foundations for measuring learning gain in higher education: A conceptual framework and measurement instrument." *Higher Education Pedagogies*, 3, 266–301, URL `https://doi.org/10.1080/23752696.2018.1484672`.

Vincent, Marin, Renaud Giess, Rémy Balthazard, Nguyen Tran, Éric Mortier, and David Joseph (2022), "Virtual aids and students' performance with haptic simulation in implantology." *Journal of Dental Education*, 86, 1015–1022.

Vincent, Marin, David Joseph, Christophe Amory, Nathalie Paoli, Pascal Ambrosini, Éric Mortier, and Nguyen Tran (2020), "Contribution of haptic simulation to analogic training environment in restorative dentistry." *Journal of Dental Education*, 84, 367–376.

Vishwanath, Dhanraj (2014), "Toward a new theory of stereopsis." *Psychological Review*, 121, 151–178.

Wages, Richard, Stefan M. Grünvogel, and Benno Grützmacher (2004), "How realistic is realism? considerations on the aesthetics of computer games." *Entertainment Computing–ICEC*, 216–225.

Walker, Judith and Hsing Chi von Bergmann (2015), "Lessons from a pilot project in cognitive task analysis: The potential role of intermediates in preclinical teaching in dental education." *Journal of Dental Education*, 79, 286–294.

Wang, D., T. Li, Y. Zhang, and J. Hou (2016a), "Survey on multisensory feedback virtual reality dental training systems." *European Journal of Dental Education*, 20, 248–260.

Wang, D. X., Y. Zhang, Y. Wang, P. Lu, and Y. Wang (2003), "Development of dental training system with haptic display." *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 159–164.

Wang, Dangxiao, Hao Tong, Youjiao Shi, and Yuru Zhang (2015a), "Interactive haptic simulation of tooth extraction by a constraint-based haptic rendering approach." *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, 278 – 284.

Wang, Dangxiao, Yuru Zhang, Jianxia Hou, Yong Wang, Peijun Lv, Yonggang Chen, and Hui Zhao (2012), "Idental: A haptic-based dental simulator and its preliminary user evaluation." *IEEE Transactions on Haptics*, 5, 332–343.

Wang, Dangxiao, Yuru Zhang, and Zhitao Sun (2007), "Multi-modal virtual reality dental training system with integrated haptic-visual-audio display." *Lecture Notes in Control and Information Sciences*, 362, 453–462.

Wang, DangXiao, YuRu Zhang, Yong Wang, PeiJun Lü, RenGe Zhou, and WanLin Zhou (2009), "Haptic rendering for dental training system." *Science in China, Series F: Information Sciences*, 52, 529–546.

Wang, Dangxiao, Siming Zhao, Teng Li, Yuru Zhang, and Xiaoyan Wang (2015b), "Preliminary evaluation of a virtual reality dental simulation system on drilling operation." *Bio-Medical Materials and Engineering*, 26, S747–S756.

Wang, Dangxiao, Xiaohan Zhao, Youjiao Shi, Yuru Zhang, Jianxia Hou, and Jing Xiao (2016b), "Six degree-of-freedom haptic simulation of probing dental caries within a narrow oral cavity." *IEEE Transactions on Haptics*, 9, 279–291, URL `http://ieeexplore.ieee.org/document/7412781/`.

Ware, Colin and Ravin Balakrishnan (1994), "Reaching for objects in vr displays." *ACM Transactions on Computer-Human Interaction*, 1, 331–356, URL `https://dl.acm.org/doi/10.1145/198425.198426`.

Wei, June and Gavriel Salvendy (2004), "The cognitive task analysis methods for job and task design: Review and reappraisal." *Behaviour and Information Technology*, 23, 273–299.

Weick, Karl E. (1990), "The vulnerable system: An analysis of the tenerife air disaster." *Journal of Management*, 16, 571–593.

Weidmann, S. M., J. A. Weatherell, and Stella M. Hamm (1967), "Variations of enamel density in sections of human teeth." *Archives of Oral Biology*, 12, 85–97.

Wertheimer, Max (1961), *Productive thinking.*, enlarged e- edition. Tavistock.

Wierinck, Els R, Veerle Puttemans, Stephan P Swinnen, and Daniel van Steenberghe (2007), "Expert performance on a virtual reality simulation system." *Journal of dental education*, 71, 759–766.

Woo, S.E., E.H. O'Boyle, and P.E. Spector (2017), "Best practices in developing, conducting, and evaluating inductive research." *Human Resource Management Review*, 27, 255–264.

Wood, David, Jerome S. Bruner, and Gail Ross (1976), "The role of tutoring in problem solving." *Journal of Child Psychology and Psychiatry*, 17, 89–100.

Wood, Harlington (1999), "United states v. lowis, 174 f.3d 881 (7th cir. 1999)." URL `https://www.courtlistener.com/opinion/763510/united-states-v-gary-lowis/`.

Wu, Jun, Dangxiao Wang, Charlie C. L. Wang, and Yuru Zhang (2010), "Toward stable and realistic haptic interaction for tooth preparation simulation." *Journal of Computing and Information Science in Engineering*, 10, 021007.

Xia, Pingjun, António Mendes Lopes, and Maria Teresa Restivo (2013), "Virtual reality and haptics for dental surgery: A personal review." *Visual Computer*, 29, 433–447.

Yamaguchi, S, Y Yoshida, H Noborio, S Murakami, and S Imazato (2013), "The usefulness of a haptic virtual reality simulator with repetitive training to teach caries removal and periodontal pocket probing skills." *Dental Materials Journal*, 32, 847–852.

Yin, Myat Su, Peter Haddawy, Siriwan Suebnukarn, Farin Kulapichitr, Phattanapon Rhienmora, Varistha Jatuwat, and Nuttanun Uthaipattanacheep (2021), "Formative feedback generation in a vr-based dental surgical skill training simulator." *Journal of Biomedical Informatics*, 114.

Yoshida, Yoshinori, Satoshi Yamaguchi, Yusuke Kawamoto, Hiroshi Noborio, Shinya Murakami, and Taiji Sohmura (2011), "Development of a multi-layered virtual tooth model for the haptic dental training system." *Dental Materials Journal*, 30, 1–6, URL `http://joi.jlc.jst.go.jp/JST.JSTAGE/dmj/2010-082?from=CrossRef`.

Zacks, Jeffrey M., Barbara Tversky, and Gowri Iyer (2001), "Perceiving, remembering, and communicating structure in events." *Journal of Experimental Psychology: General*, 130, 29–58.

Zafar, Sobia, Yvonne Lai, Christopher Sexton, and Allauddin Siddiqi (2020), "Virtual reality as a novel educational tool in pre-clinical paediatric dentistry training: Students' perceptions." *International Journal of Paediatric Dentistry*, 30, 791–797.

Zheng, Fei, Wen Feng Lu, Yoke San Wong, and Kelvin Weng Chiong Foong (2013), "Graphic processing units (gpus)-based haptic simulator for dental implant surgery." *Journal of Computing and Information Science in Engineering*, 13, 041005, URL `http://computingengineering.asmedigitalcollection.asme.org/article.aspx?doi=10.1115/1.4024972`.

Ziane-Casenave, Sophia, Marthe Mauroux, Raphaël Devillard, and Olivia Kérourédan (2022), "Influence of practical and clinical experience on dexterity performance measured using haptic virtual reality simulator." *European Journal of Dental Education*, 26, 838–848.

Zigmont, Jason J, Liana J Kappus, and Stephanie N Sudikoff (2011), "Theoretical foundations of learning through simulation." *YSPER*, 35, 47–51, URL `http://dx.doi.org/10.1053/j.semperi.2011.01.002`.

Appendices

Appendix A

Publications

Below is a list of publications and presentations that were contributed to or derived from the work presented in this thesis:

Towers, A., Maddock, S., Stokes, C., Field, J., & Martin, N. (2018). The use and application of Virtual Reality in Dental Education [Conference presentation]. School of Clinical Dentistry Research Day, Sheffield, United Kingdom

Towers, A., Dixon, J., Martin, N. (2018, August 22-24) Pre-treatment simulation of patient-specific operative procedures [Conference presentation]. The Association for Dental Education in Europe Annual Meeting, Oslo, Norway

Vital, S., Dixon, J., **Towers, A.**, Field, J. (2019, August 21-23) Categorising learning outcomes for operative pre-clinical skills [SIG co-chair]. The Association for Dental Education in Europe Annual Meeting, Berlin, Germany

Towers, A., Dixon, J., Martin, R., Field, J., Martin, N. (2019, August 21-23) Virtual Reality Operative Dentistry at The University of Sheffield [TechExpo Presentation]. The Association for Dental Education in Europe Annual Meeting, Berlin, Germany

Towers, A., Field, J., Stokes, C., Maddock, S., & Martin, N. (2019). A scoping review of the use and application of virtual reality in pre-clinical dental education. British Dental Journal, 226(5), 358–366. https://doi.org/10.1038/s41415-019-0041-0

Field, J., **Towers, A.**, Dixon, J. (2020, November 4) Continuing to Bridge the Clinical Gap [Symposium presentation]. The Association for Dental Education in Europe, Digital Dentistry Online Symposium

Dixon, J., **Towers, A.**, Martin, N., & Field, J. (2021). Re-defining the virtual reality dental simulator: Demonstrating concurrent validity of clinically relevant assessment and

feedback. European Journal of Dental Education, 25(1). https://doi.org/10.1111/eje.12581

Field, J., Dixon, J., **Towers, A.**, Green, R., Albagami, H., Lambourn, G., Mallinson, J., Fokkinga, W., Tricio-Pesce, J., Crnić, T., & Vital, S. (2021). Defining dental operative skills curricula: An ADEE consensus paper. European Journal of Dental Education, 25(2), 405–414. https://doi.org/10.1111/eje.12595

Towers, A., Dixon, J., Field, J., Martin, R., & Martin, N. (2022). Combining virtual reality and 3D-printed models to simulate patient-specific dental operative procedures – A study exploring student perceptions. European Journal of Dental Education, 26(2), 393–403. https://doi.org/10.1111/eje.12715

Towers, A., Maddock, S., Stokes, C., Field, J., & Martin, N. (2023). Task Analysis of Caries Removal and Cavity Preparation Using Virtual Reality [Conference presentation]. School of Clinical Dentistry Research Day, Sheffield, United Kingdom

al Ali, H., Nassief, S., **Towers, A.**, Field, J., & Martin, N. (2024). The value of stereoscopic three-dimensional vision on dental students' performance in a virtual reality simulator. Journal of Dental Education. https://doi.org/10.1002/jdd.13630

Nassief, S., al Ali, H., **Towers, A.**, Field, J., & Martin, N. (2024). Dental students' perceptions of the use of two-dimensional and three-dimensional vision in dental education using a three-dimensional haptic simulator: A qualitative study. Journal of Dental Education. https://doi.org/10.1002/jdd.13682

Towers, A., Stokes, C., Maddock, S., & Martin, N. (2024, September 7-10) Task Analysis for Operative Task Design Using Virtual Reality [Conference presentation]. The Association for Dental Education in Europe Annual Meeting, Leuven, Belgium

Sarstedt, M., Martin, N., **Towers, A.** (2024, September 7-10) VR Simulation… choices were made [Conference presentation]. The Association for Dental Education in Europe Annual Meeting, Leuven, Belgium

Appendix B

Review of Task Analysis, Information Gathering and Knowledge Elicitation Methods

This appendix details the information gathering and task analysis methods that were evaluated and/or contributed to the method used in the study presented in Chapter 5.

B.1 Methods of Information Gathering and Knowledge Elicitation

Knowledge elicitation techniques are the tools for information gathering that is fed in to the task analysis approaches. Some elicitation techniques are more appropriate than others for a given task analysis approach, however a task analyst can draw on any of the techniques to meet the goals of the task analysis at hand.

B.1.1 Documentation Analysis

Document analysis is likely to be the most widely used method for collecting data (Jonassen et al., 1999, p. 237). It involves consulting documentation, user guides, standards, manuals and so on in order gain information about tasks including the procedures used, performance standards, underlying concepts and known problems. It is rarely the only source of information for a task analysis but it is often used to prepare for the analysis itself so that time isn't wasted obtaining information that is already readily available (Kirwan and Ainsworth, 1992, p. 35) or to triangulate information gathered from other sources and verify the correctness of descriptions. (Jonassen et al., 1999, p. 237)

B.1.2 Questionnaires

A widely used tool for gathering information in a fast and efficient way is to conduct a survey using a questionnaire. They allow for efficient administration because they can be answered simultaneously by a number of people and online questionnaires permit the automatic collation of the responses. This allows for a breadth of opinion to be gathered from a large number of experts and compared (Kirwan and Ainsworth, 1992, p. 64). They can also be used in conjunction with other approaches, for example to validate information gathered from observing a small number of performers with a larger cohort of respondents (Jonassen et al., 1999, p. 247).

With the goal of deconstructing tasks, a questionnaire may not be the most suitable approach as it places onus on the respondent to imagine a situation and then comment upon the steps. This may lead respondents to be influenced by what they believe is required from their answer or misunderstand what a question is asking through subjective interpretation (Kirwan and Ainsworth, 1992, p. 65). Designing an effective questionnaire always presents challenge (Jonassen et al., 1999, p. 248) and it is questionable if this is an appropriate method to capture a description of a psychomotor skill. However, the approach may be useful in validating the deconstructed model once it has been established.

B.1.3 Observation

Observational techniques aim to obtain data by observing how a task is normally performed. The approach can trace its roots back to early time and motion studies and is arguably the most valid method of collecting task analysis data because any resultant training material will be based on the real-world actions of a competent performer (Jonassen et al., 1999, p. 241). A number of observational techniques are available including: direct visual observation, remote live visual observation (e.g. via video-link), video recording, time-lapse photography etc. as appropriate to the requirements of the study (Kirwan and Ainsworth, 1992, p. 53).

Observations can be classified as Obtrusive and Unobtrusive (Jonassen et al., 1999, p. 242). An unobtrusive observation intends to capture the tasks exactly as performed with the observer endeavouring to not alter the way the task is usually performed (endeavouring to avoid a Hawthorn effect). Kirwan and Ainsworth (1992, p. 55) defines 3 levels of intrusion when observing: The lowest level of these is classed as *observer unobserved* and uses, for example, a video link to allow remote observation. The next level of intrusion is *observer observed* where the observer is co-located and the operator is aware they are being observed. Finally, *observer participant* has the observer actively taking part alongside the operator, which is useful when observation of the dynamics of team performance are being investigated.

Whilst observational techniques can provide information that cannot be acquired in any other way, the information is captured in a very 'raw' form and can require considerable analysis to make sense of the data (Kirwan and Ainsworth, 1992, p. 57). Furthermore, when the task requires mental operations that inform decisions, an unobtrusive observation cannot offer any insight in to the underlying thought processes. The experts actions and motions can be observed but not the reasons behind why they made those particular motions (Kirwan and Ainsworth, 1992, p. 58), (Jonassen et al., 1999, p. 242). In the context of this work an unobtrusive observation will not be sufficient to reveal the required cognitive aspects underlying the performance.

B.1.4 Interviews

Individual interviews are one of the most common tools used in task analysis. They permit the analyst to question an expert about the task being analysed which sometimes provides an adequate description of the task alone or provides insight in to the kind of information that will be required using other means (Jonassen et al., 1999, p. 253). The interview can gather information beyond simply how the task is performed and can include related misconceptions, common errors and best-practice approaches in a much shorter time than many other approaches. However, in many cases the success of the approach is predicated on a well conducted interview by a competent interviewer and the interviewee being an expert (Jonassen et al., 1999, p. 253) although in some cases a novice can be interviewed to identify areas of difficulty (Kirwan and Ainsworth, 1992, p. 66).

Interviews can be categorised as unstructured or structured. Unstructured interviews do not set a firm agenda of questions or problems to solve. They permit a more spontaneous and flexible exploration of a wider range of ideas and problems. In a task analysis context, this kind of interview is most appropriate at the beginning of the analysis and can provide structure for subsequent investigation and targeted structured interviews. However, care must be taken to avoid the unstructured formatting resulting in a loss of focus and giving too much importance to the individual biases or problems of the interviewee (Jonassen et al., 1999, p. 254).

Structured interviews are more directed and use a preplanned agenda of questions and topics. Preparing for a structured interview requires the task analyst to have a stronger background in the subject area to pose appropriate questions (Jonassen et al., 1999, p. 254). However, this structure allows for a more efficient use of time and guide the responses towards areas of interest, with the trade-off being a discouragement of spontaneous exploration of ideas that emerge in the interview. It is good practice to end an interview with an open ended question to ask if anything has been missed that the analyst should be aware of (Kirwan and Ainsworth, 1992, p. 67)

B.1.5 Unstructured group interviews: focus groups and brainstorming

Focus groups and brainstorming are a relatively unstructured, but simple, method to gather the reaction of a group of people and gather experts to discuss all of the aspects of a problem(Kirwan and Ainsworth, 1992, p. 157). The loose structure permits participants to introduce new ideas, improvements and suggestions hence their popularity in marketing and product development where a group's perception of a new product, feature or service can be gathered with relative ease (Jonassen et al., 1999, p. 263). Brainstorming is a technique that can be used as a tool for problem solving; sessions often commence by setting the scene or presenting a problem and tasking participants with coming up with solutions to the problems. Brainstorming is at its most effective when attempting to find alternative solutions to problems, creativity is encouraged when suggesting ideas and (other than clarification) discussion around the technical details of how practical an idea is to solve the problem is discussed in a subsequent stage (Furnham, 2000).

Although these techniques were not specifically developed for task analysis, they permit information about different ways of accomplishing a task to be discussed, and are particularly useful for tasks involving mental operations (Jonassen et al., 1999, p. 263). They can lead to better judgements than would be made by an individual but group dynamics and disagreements can impact upon the process this could steer the outcomes from the discussion towards the opinion of a particularly forceful or senior participant (Kirwan and Ainsworth, 1992, p. 157) .

B.1.6 Structured group interviews: Delphi technique

The Delphi technique is a structured group interview technique that helps establish a consensus about a particular issue. Experts are first asked to individually provide a judgement on the area of interest. Responses are then tabulated then the experts are given the opportunity to revise their response in the light of the (anonymous) responses provided by other experts (Kirwan and Ainsworth, 1992, p. 158). After several rounds of questioning an average response can be taken as the best group estimate.

Whilst not prohibited, the Delphi technique does not tend to assemble the group together, nor do the participants usually meet each other, rather they provide answers to questionnaires then privately revise their own answers in the light of the tabulated responses of the group. This reduces the influence of dominant members of the group and, over time, results in a convergence of opinion (Jonassen et al., 1999, p. 269). However, the reliance on questionnaires means that it is not as suitable for describing psychomotor skills.

B.1.7 Think-aloud Protocols

Think-aloud Protocols are a combination of interviewing and observation. Whilst carrying out a task, an expert provides a verbal narrative of what they are thinking and decisions they are making. However, the verbal description can also reveal the strategies employed, observations, how problems were overcome or even comments and insight into the emotional reactions encountered during the task (Jonassen et al., 1999, p. 259).

The result of this protocol is a model of the problem solving performance, including the invisible thought processes that informed what was observed. This makes it suitable for overt motor-skills where observation alone cannot provide insight into the decisions that guide the observable actions (Jonassen et al., 1999, p. 260). However, think-aloud is not an appropriate method in some contexts; it could have ethical implications, for example it could not be used for investigating the cognitive processes used by Human Resources staff when terminating an employee. Also, during an operative procedure it would not be appropriate for the surgeon's attention to be diverted to providing a narrative (especially if that narrative may cause a conscious patient alarm or distress). In these situations, simulation or role-play can be used or a retrospective protocol where a performer provides a narrative to a previously encountered situation and attempts to describe the decisions and reactions encountered at the time. The use of simulation can confound the investigation because the task is artificial, but retrospective narratives can introduce memory lapses and bias (Jonassen et al., 1999, p. 260) and only verbalisations that rapidly follow a thought process can be taken as reflective of the conscious thoughts of the task (Charters, 2003).

The think-aloud approach reveals a lot of information in a short period of time and due to observing the actual performance, the results have high face validity (Jonassen et al., 1999, p. 262). However, thinking aloud is awkward for many performers and can interfere with thinking (Jonassen et al., 1999, p. 262) and verbalising physical actions can be distort the process when meeting the demands of a think-aloud task (Charters, 2003) because other processes crowd out verbal information from working memory (Ericsson and Simon, 1980). Some participants may require coaching on how to think-aloud or explicit guidance. However, this can risk 'leading the witness' and further interfere with the normal thought process used in the execution of the task (Charters, 2003).

Much of our thought is not held in a verbal form and as knowledge is internalised it becomes increasingly abstract with words only forming part of the meaning (Charters, 2003). This means that making sense of a transcript taken from a think-aloud protocol will be more challenging than normal speech or writing, because the purpose of these, often incomplete, inner vocalisations is not meant to be communicated to anyone other than the thinker (Charters, 2003). Consequently, a think-aloud protocol may only reveal part of the problem-solving skill relied upon by the performer. Deeper knowledge is not always revealed because the true complexity has to be simplified in to words before even the thinker themselves is aware of it (Charters, 2003). This means that the participant may not be consciously aware of (or have even noticed) that they are relying on knowledge and, because working memory is finite, any realisation may be lost before it is verbally expressed as new thoughts supersede it (Kirwan and Ainsworth, 1992; Jonassen et al., 1999; Charters, 2003).

The think-aloud technique is one of the most effective methods of accessing higher-level thinking providing the most direct means available to access information about the mental processes used (Kirwan and Ainsworth, 1992, p. 77). It can be used to investigate individual differences when performing the same task, permitting each participant to be viewed as a mini-case study (Olson et al., 1994). Whilst verbalisations are often incomplete, what is present is a reliable source of insight in to the thought processes and has been demonstrated across many experimental settings (Ericsson and Simon, 1980). These verbalised accounts should be triangulated with task data captured from other approaches (Kirwan and Ainsworth, 1992, p. 71), supplemented with questioning for additional detail (Kirwan and Ainsworth, 1992, p. 74) and retrospective questioning to allow the participant to expand and add depth about their thought processes (Charters, 2003)

B.2 Task Analysis Methods

There are many approaches to task analysis, 24 of which were identified in Jonassen et al. (1999) as having particular applicability to instructional design. These approaches form the core of the exploration of candidate approaches, supplemented by additional methods identified in a broader search. However, within these 24, there are a number that are conceded as philosophical frameworks providing a way of thinking about tasks and instructional design rather than structured approaches to task analysis. Whilst of interest, the practical focus of this project requires a stronger approach so these will be noted but not explored further.

This section provides an overview of each of the task analysis methods considered and discuss any literature found using the broad search discussed in 5.3. In summary Web Of Science™was queried using the following criteria:

Any Field: *Name of method* AND Any Field: Dent*

B.2.1 Task Description

Task description attempts to describe the interactions between the performer, the equipment used and the wider task context. It views a system as a collection of interrelated parts and that the overall success depends on success within the subsystems (Jonassen et al., 1999, p. 35). This approach starts with a view of the overall system and then works its way down to specify what individuals must do with the equipment to achieve the system goals. The approach has two major steps *task description* and *task analysis.* The task description is a detailed description of the context of the task including what triggers action and the proper response to a given stimulus. Generally, the description begins more generally with an emphasis on observable behaviour and then becomes more specific (Jonassen et al., 1999, p. 37) breaking the steps down in to increasing detail. In the task analysis phase, the analyst considers tasks that underlie the capabilities essential for good performance and then examines the description in order to find a better way of performing the task which, it is assumed, the analyst can identify through rational thought rather than in consultation with subject matter experts (Jonassen et al., 1999, p. 40-41).

The wider system-level focus suggests that this approach initially takes too broad a view for this project. In the context of the functionality that the Virteasy simulator provides, dedicating analysis-time to the wider context of the task rather than the specific steps within the procedure will not result in useful information. Additionally, shifting the onus of identifying the capabilities associated with good performance to the task analyst by not including subject expert input in a dental operative task does not seem appropriate.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.2 Procedural Analysis

Procedural analysis is used to describe the job performance of labourers and skilled workers. It has seen use in describing assembly, service and repair tasks as a series of discrete actions with its primary function is describing job tasks and motor skills tasks, but can be used for cognitive activities as long as those steps can be described as observable performances (Jonassen et al., 1999, p. 45). Procedural analysis describes task performance as a linear series of steps, which are best represented as a series of predominantly observable behaviours. The analyst observes a skilled performer and produces a step-by-step list of the actions observed and attempts to capture any branching conditions to handle decisions made by the operator. This step by step list is then compiled into a procedural flowchart (optionally in collaboration with the subject matter expert) and verified that the description is complete.

The focus on phrasing tasks as observable behaviours so that performance is evaluated in terms of completing an overt sequence of actions suggests that this approach will not result in a description of the task with a focus on the underlying cognitive elements. Whilst caries removal and cavity preparation are overt actions, the decision making behind those actions is of key importance to this study. This approach is not as suited for uncovering this covert knowledge (Jonassen et al., 1999, p. 53) therefore it is considered inappropriate for this work.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.3 Job Task Analysis

Job Task Analysis was developed as a process for producing vocational instruction. It is rooted in a behavioural framework where the emphasis is placed on what a person *does* and not what they know. The observable behaviour sets the goals for the training (Jonassen et al., 1999, p. 55). The general approach of deriving and describing learning objectives, developing training to meet those objectives and then determining how well those objectives are achieved appears to be compatible with the goals of this project. However, the method discounts the cognitive aspects and is considered inappropriate for these broader educational goals (Jonassen et al., 1999, p. 62) is unlikely to lead to a useful description of the task.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.4 Functional Job Analysis

Functional Job Analysis focusses directly on what employees do to achieve a task, describing what is done in specific terms rather than the overall goals. A task analyst using this approach should ask why each task is done and how it contributes to the organisation's goals; providing a consistent definition of the complexity and involvement of an individual in the execution of tasks (Jonassen et al., 1999, p. 63). Functional Job Analysis defines each task in 5 components:

- Who carries out the task?
- What is done?
- Why is the task carried out?
- What tools, equipment etc. are necessary?

- What instructions are followed?

The theory states that all jobs can be described in terms of one of three basic concepts, therefore activities are further divided in to hierarchies of *data, people* and *things*. The origin of this approach is to give employers a tool with which to define a person's job and it dedicates much emphasis on the flow of data and information between people and their interactions within the organisation. These 3 categorisations are very coarse and are indicative of the level that this approach is intended to operate at - primarily inventorying tasks (Jonassen et al., 1999, p. 72). Therefore, the approach is not considered suitable for the definition of a pre-selected operative procedure at a sufficient level of detail to support the goals of the project.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.5 Learning Hierarchy (Prerequisites) Analysis

A Learning Hierarchy describes how a learner can solve a problem by representing the pre-requisite skills necessary to do so. The structure, often presented graphically, presents a hierarchical relationship between skills with simple tasks at the bottom and complex at the top. This indicates which lower-order goals and skills must be learned in order to master the higher level tasks. Because the lower level tasks are pre-requisite to the higher, the approach is often referred to as pre-requisites analysis. The approach was introduced in Gagne (1962) and stated that there was a hierarchy of skills and that the best predictor of an individual's mastery of any given skill would be their mastery of the prerequisite skills. Instructional designers often begin developing material by first constructing a hierarchy to suggest the order that instruction must be provided in (Jonassen et al., 1999, p. 77). Any time that a a learning task requires an intellectual skill such as discrimination, application of a rule or concept; a learning hierarchy can be used (Jonassen et al., 1999, p. 78).

Learning hierarchy analysis assumes that the hierarchy of skills can be discovered through rational analysis. When arranged in this way, this order represents the ideal teaching sequence because learning is assumed to be cumulative (Gagné and Medsker, 1996). However, Berger et al. (1980) has criticised the completeness of this bottom-up sequence because no sequence or relationship is provided between co-ordinate skills at the same level in the hierarchy nor is it specified how much dependence a super-ordinate task places on each of its subordinates.

Learning hierarchy analysis is a proven technique for organising instruction and has been used in a number of subject areas (Jonassen et al., 1999, p. 83) and because the hierarchies indicate an instructional sequence, this facilitates transfer from simple to more complex skills. This approach appears to fit well with the goals of this project, however, it may be more suited to organising the instruction at a higher subject-level than capturing the underlying knowledge of a specific task. Within a specific task, the hierarchy in terms of 'learning' is likely to be quite flat and much of the lower level falling outside of the scope of the simulation experience. Furthermore, it has been questioned if learning outcomes such as the development of mental models can be taught in a bottom-up fashion (Jonassen et al., 1999, p. 85) such as this.

Because this technique is known under two names, a search of the literature was performed using the above criteria for both "Learning Hierarchy" and "Prerequisites Analysis". No matches were found using either name indicating this approach has not been used in a dental context at the time of writing.

B.2.6 Hierarchical Task Analysis

Hierarchical Task Analysis (HTA) shares similarities with Learning Hierarchy Analysis in that it also seeks to describe tasks in a hierarchical form. However, it it does so by what was at the time, a radically new approach of task description based on functional rather than behavioural/psychometric constructs of the task (Annett, 2004). As a method, it is best viewed as a generic approach to the investigation of problems of human performance in a goal-directed context rather than as a strict procedure (Annett, 2004). It was developed fro a need to understand the complex non-repetitive skills required in process control situations such as those found in steel production, chemical/petrol refining and power generation. Prior theories were based on the notion that complex behaviour comprised of a sequence of stimulus-response connections and that a teacher had to identify these individual components and reinforce each one individually to produce progressively closer approximations of the the desired behaviour (Annett, 2004).

HTA begins by hierarchically decomposing the goals (the end state) and subgoals before considering the actions required to accomplish each goal (Salmon et al., 2010), it may be used alone or incorporated with additional task analyses (French et al., 2019). This functional approach is taken from systems theory and information processing models of human performance which define task performance in terms of the interaction between humans and machine (Annett, 2004, CHECK ME) and focusses on what an operator is required to do in terms of actions and cognitive processes to achieve a goal (Salmon et al., 2010). The end result of the HTA is a detailed top-down description of the task based around the concept of a *plan* and the *operations* that contribute to it (French et al., 2019; Annett, 2004). This allows for a fully documented training programme to be produced at the level of detail needed to produce very specific how-to guides in plain language at a level appropriate to the purpose of the analysis (Annett, 2004). Representing an activity using a hierarchy such as this is valuable in its own right and the usefulness and flexibility of the outputs mean this is a very popular approach

(Salmon et al., 2010).

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.7 Task Decomposition

Task Decomposition is seen as a sub-step of Hierarchical Task Analysis by Annett (2004) but as a technique in its own right that builds on a previously conducted HTA by Kirwan and Ainsworth (1992). However, HTA can be used in conjunction with other methods (French et al., 2019) so it is to be expected that the lines between complimentary approaches may blur. Task decomposition is a structured approach to expanding information gathered from a task description to focus on areas of specific interest. It determines the psychological factors essential to successful performance, describes them in such a way as to facilitate training for the task and does so in such a way that a sound knowledge of results and performance can be provided to the learner (Miller, 1953). Subjects trained using materials derived from this approach outperform control groups and the skills are transferable to other domains (Frederiksen and White, 1989). The method assists the analyst in uncovering this information by reminding them where to focus their investigations towards the discriminations, decisions and responses that are important at each step in the execution of the task. This ensures that the analysis is reliable, provides a consistent level of description, in a formally stated conceptual structure with directed and specific format for the analysis (Miller, 1953).

Task Decomposition reminds the analyst that a task has a goal, but goes further than HTA by stating that the better the task goal is defined the better it can serve as a criterion for measuring performance effectiveness (Miller, 1953). It is preferable to make a *gross* task analysis first and then complete the description of the behaviour when analysing the sub tasks. The decomposition of the task can follow a number of categories, such as the cues initiating action, decisions that must be made, typical errors, criteria for determining success. Kirwan and Ainsworth (1992, p. 97) provides a comprehensive list, elaborating upon Miller's original categories but these can be customised by an analyst to focus on areas of specific interest. As tasks are decomposed in this manner, skills may be identified that are not observable in expert performance but are necessary precursors to the acquisition of the skill as demonstrated by an expert (Frederiksen and White, 1989).

Task decomposition can require a significant amount of an analyst's time to cross-check information sources but the structure that the approach provides ensures that the areas of interest are systematically considered leading to a detailed description of the task (Kirwan and Ainsworth, 1992, p. 104).

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.8 Information Processing Analysis

Information Processing Analysis (IPA) is a method that can reveal the cognitive operations and decisions necessary to achieve a task. Whereas other methods identify overt behaviours, IPA focusses on describing the internal mental steps or operations required by an operator to perform a task (Jonassen et al., 1999, p. 87). IPA describes cognitive task performance as a series of operations and decisions with a determined beginning and end. Performance of a task is triggered by the need to act either through the recognition of a problem or receipt of instructions and ends with the task's completion or abandonment. IPA is usually applied to higher level cognitive tasks (Jonassen et al., 1999, p. 88) and recognises that an empirically derived IPA may not generate an idealised description of a task because different performers will describe different ways of achieving the same outcome. This will be particularly noticeable across different levels of expertise (Jonassen et al., 1999, p. 89). IPA is not the recommended approach for tasks involving observable performance (Jonassen et al., 1999, p. 89) and produces a comprehensive description of what is done rather than breaking the tasks down in to any kind of hierarchy. As the task to be analysed is predominantly an observable performance and a desirable output is a deconstruction of the task (rather than a description of it) this approach is not considered appropriate for this task.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

Learning Contingency Analysis

Learning Contingency Analysis focusses on tasks within a learning environment rather than those associated with job performance in order to identify the behavioural components of tasks (Jonassen et al., 1999, p. 99). The approach determines interdependence among the components in order to sequence learning activities and determine instructional strategies to teach those tasks. The approach deals with two aspects: sequencing of content and determining the conditions under which the content is best taught. By examining the interrelations between content the analyst is able to understand how learning one item of content will enable the learning of another. The origins of the approach can be traced to approaches from behavioural psychology (Jonassen et al., 1999, p. 99) and in order to build complex behaviour the skills must be broken down and trained in such a fashion that one item leads to the next. As a behaviourist model, tasks are identified and then behaviours that lead to them are identified. These are then sequenced so that they build upon each other to teach the skill. However, this project is not behaviourist in nature. It is based on the premise that rules and underlying cognitive knowledge determine the actions rather than a sequential linear task. An approach that does not deal with mental processing or non-directly observable behaviours (Jonassen et al., 1999, p. 105) is not appropriate for this task.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

B.2.9 Cognitive task analysis

Cognitive Task Analysis (CTA) is a family of methods for describing the underlying knowledge and thinking that is used when performing a task (Jonassen et al., 1999; Walker and von Bergmann, 2015). Traditional task analysis emphasises the desired target performance, whereas CTA attempts to address the knowledge base for the whole job including its organisation and connections between knowledge elements (Wei and Salvendy, 2004). Not all CTA methods are useful for conducting a task analysis for instructional design (Jonassen et al., 1999, p 107) so discussion of the methods will be restricted to those that have been identified as useful. CTA for instruction includes a description of the actions and the decisions that must be made to perform them (Jonassen et al., 1999, p 107). Many of the unobservable components of task performance are mental activities such as decision making and problem solving; these are good candidates for CTA (Wei and Salvendy, 2004).

GOMS Analysis

GOMS is a CTA method for analysing and modelling the knowledge and skills required to perform tasks using a device (Jonassen et al., 1999, p 111)(Jonassen et al., 1999, p 107). It's name is an acronym for "Goals-Operators-Methods-Selection" and is particularly suited for goal-orientated tasks in a computer environment. It describes each task-component in terms of the goals and sub-goals that the user hopes to accomplish as as an action-object pair. Next, methods consisting of low-level repetitive operations are described that will be used to accomplish the goals. Often, there is more than one method to achieve a goal, so selection rules describe how to select amongst the valid options. The primary application of GOMS has been for interface design and profiling the interactions with different interface implementations (Jonassen et al., 1999, p 117). Whilst researchers agree that GOMS can be used for conducting analysis of complex goals it is generally argued that this is not advisable (Jonassen et al., 1999, p 118).

A search of the literature using the criteria above resulted in one match, but this was not in the context of a task analysis for instructional design. The same result was returned for both the acronym and full definition.

PARI

The Precursor-Action-Results-Interpretation (PARI) method is intended to analyse the system, procedural and strategic knowledge required for troubleshooting problems in real-world settings, with a focus on the actions that are taken whilst solving problems (Jonassen et al., 1999, p 121). It attempts to identify each decision made before an action, record the actions taken, their results and the expert's interpretation of them. This information enables a knowledge base of trouble-shooting guidance to be produced so that similar problems can be solved in the future (Jonassen et al., 1999, p 121). This approach has seen the most success when applied to complex systems such as in aerospace where trouble-shooting guides were developed to support maintenance staff (Jonassen et al., 1999, p 121).

In the context of the goals of this project, whilst the operative task could be conceptualised as a problem-solving activity, it does not have the unpredictability and inconsistency about which concepts or rules should be applied that this method attempts to encapsulate. Fundamentally, the selected operative task is a psychomotor skill and requires some effort to be construed as a trouble-shooting task in order fit this method.

Using the search criteria above, PARI found 56 matches and zero matches when the acronym was expanded. The 56 results were examined and it was found that these were all false positives, for example matching on the author name Pari; part of a place name e.g. the *Pari Aike Formation of Southern Patagonia*; the term *pari passu* and so on. No results were found for the use of PARI as a task analysis method in a dental context.

DNA

The CTA method of Decompose-Network-Assess is an approach to elicit skills and knowledge information from experts. It identifies a hierarchy of structured knowledge and curriculum information aimed at producing training information and a model of expert performance for use in intelligent systems (Jonassen et al., 1999, p 131). It attempts to aid the analysis process and improve the efficiency by being implemented in a series of interactive computer programs that are used by the parties involved in the analysis to build up a database of task information (Jonassen et al., 1999, p 131). The goal for this project is not to produce an expert system therefore this approach is not considered appropriate.

For obvious reasons, the frequently used term DNA resulted in tens of thousands of results when applied to the broad search criteria detailed above, but when And'ed with "task analysis" returned zero results. The expanded acronym also resulted in zero matches.

Cognitive Simulations

Cognitive simulations should be built to articulate and understand theoretical ideas. They encapsulate mental constructs as functioning computer programs, with early cognitive simulations realised as LISP or Prolog programs (Jonassen et al., 1999, p 139). These simulations are a formal representation of the reasoning applied by experts to a given situation. The goal of this project is not to create an expert system so this approach is not appropriate for the task analysis in this case.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing.

Case Based Reasoning

Knowledge is often represented in the form of stories so analysis of the knowledge of tasks can be performed through investigating these stories. This approach is known as case based reasoning. Stories can be recalled to assist with diagnosing problems, hypothesising solutions and the design of applications (Jonassen et al., 1999, p 147). The subject matter expert is asked to recall solutions to previously-encountered problems and how they were solved. These are then applied to the current situation to solve the current problem. The approach was developed to design and build artificial intelligence systems and is not commonly used for task analysis, but when the process is applied to develop intelligent systems, this process is effectively a task analysis (Jonassen et al., 1999, p 148). The concept of knowledge being represented as stories is appealing and a search using the criteria above returned 15 results, many using the approach in a dental context (albeit not for task analysis). However, this approach may be more appropriate to deal with non-standard unique and interesting cases where the story holds more meaning than for use describing the basics of a core operative skill. Additionally, the story based content may suffer from the same analogical transfer effects as discussed in previous sections which could limit its effectiveness. Therefore, this approach is not considered appropriate for this study.

Cognitive task analysis

As noted above, CTA is a family of methods, however, it appears to have established an identity in its own right in the literature detached from the 'family' of methods such as those described in Jonassen et al. (1999). Many publications were found to state that CTA was used but did not specify a specific 'family member' as used for the analysis. Therefore, to ensure relevant publications are not missed an additional search for 'Cognitive task analysis' was performed. Using the criteria above, this returned 10 results, 2 of which were considered relevant for detailed review. Of the results that were discarded: 4

were from non-dental domains. Suebnukarn et al. (2013) and Thyvalikakath et al. (2014) used CTA to enhance the usability of electronic health record systems; Mislevy et al. (1999) used CTA to uncover the underlying skills and knowledge in treatment planning and monitoring and Suebnukarn et al. (2015) used CTA to uncover the information processing techniques used by experts and novices for diagnosis and planning of treatment and the preparation for the procedure. Whilst interesting these studies did not focus on the development of a psychomotor skill so could only provide interesting wider reading. The remaining 2 were considered relevant and investigated further.

Walker and von Bergmann (2015) used CTA to identify steps within Class II wax carving that present the highest cognitive load. The motivation for this was to deconstruct the procedure in to separate steps and then develop improved and focussed teaching tools for teaching novices the steps with highest cognitive load. For the CTA, the authors used qualified practitioners and 2nd year students trained in the task as "intermediates", so called due to their skill levels being between that of novices and experts. Following a 'think-aloud' protocol they inferred that any unduly long pauses in the narration could be associated with an increase in cognitive load that interfered with the ability to describe what was happening. From this study they found that CTA uncovered the invisible knowledge used by task performers and that the intermediates provided a valuable comparison in the identification of high-load areas because of 'expert blind spots' where the expert was not aware that a task was difficult. Overall, they recommended the use of task analysis approaches to identify areas where dental students struggle and as a means to produce improved learning material.

B.2.10 Activity Theory

Activity Theory is deployed in the context that people perform activities in. Rather than focussing on knowledge, it directs attention to the activities engaged in, the tools they use and the social context and interpersonal relationships with collaborators (Jonassen et al., 1999, p. 159). Traditional approaches neglect considerations of the context and wider environment, but Activity Theory argues that learning activities cannot be detached from the context that they will be used in. This is because the environment itself acts upon the learner to support or hinder learning so learning must be considered in its socio-cultural context (Jonassen et al., 1999, p. 160). This agrees with much of the work in specificity, where the context is part of the learner's representation of the skill so when the skill is deployed elsewhere with different feedback the performance is reduced (Proteau, 1992).

Activity Theory does not provide a prescriptive methodology for task analysis, rather providing a "philosophical framework" (Jonassen et al., 1999, p. 159) or a different way of thinking about and analysing the process of designing instructional materials. But it is argued that use of the theory increases the probability of success because the instruction takes into account because critical factors of, and the dynamics of, the environment will have been considered (Jonassen et al., 1999, p. 159). The use of VR for training is interesting in the context of Activity Theory. By definition, a simulation is not situated in the context of the environment but it does attempt to simulate it. However, in the context of this project to develop training approaches for use in a VR environment, the VR simulator itself becomes the context and therefore is situated.

One of Activity Theory's strengths is the contextual meaning and historical aspects of the task, but given this project is focussing on cognitive aspects of the task and the approach is stated to produce an overwhelming amount of information (Jonassen et al., 1999, p. 171) it is not likely to produce the desired analysis if applied directly in this case.

A search of the literature using the criteria above resulted in 10 matches, the abstracts for these were reviewed manually. None of the published works used Activity Theory as a tool for task analysis, but after excluding erroneous search results, it has seen application in a dental education context to: teach research and academic writing at undergraduate level to first (Esambe et al., 2016) and final year (Esambe, 2018) dental technology students; as a mechanism for evaluating and improving student clinical placements (O'Keefe et al., 2016) and clinical supervision (O'Keefe et al., 2014); as an approach to measure the impact of instructional videos for Dental Hygiene students' clinical skills acquisition (Lockwood et al., 2018) and as an exploration of the boundary-crossing instructional actions in inter-professional oral health care (Teras, 2016).

B.2.11 Critical incident/critical decision method

Critical decision method can be used to identify the most critical elements of a job or skill. The approach isolates and prioritises the behaviours that are essential to a job through the collection of real-world data based on specific non-routine events that challenge a practitioner's expertise. The information is retrieved via semi-structured interviews with specific probing to elicit the goals, operational and contextual cues and situational assessments led to specific conclusions and resultant actions (Jonassen et al., 1999, p. 181).

The approach has seen wide use within a dental context. Using the search format described above, the term "Critical Incident" returned 31 results . However, these studies did not use the approach in a task analysis context, rather, it was used for evaluation and retrospective insight into events that occurred in a clinical or educational setting. Whilst appealing, this approach is based around a subject matter expert identifying non-routine (and/or especially challenging) cases where the differences between the performance of a novice and expert would be particularly pronounced. In the context of this project, where a description of an operative procedure is required to provide the early educational material to a novice, this approach is not suitable. However, future work may find value in using this approach to identify challenging cases and then re-apply the approach discussed below in order to create new exercises that can prepare learners for non-routine eventualities.

The alternative term for this approach, "Critical Decision", was also searched for. This returned 14 results, however, a manual review of the abstracts showed that the match was from a descriptive natural language use of the term rather than in reference to the Critical Decision method.

B.2.12 Task knowledge structures

Task Knowledge Structures (TKS) are developed in order to identify the knowledge structures skilled performers use when performing tasks. They were originally developed in order to analyse task structures to design human-computer interactions (Jonassen et al., 1999, p. 193). TKS represent complex activities as high level concepts and are based on the assumption that all activities are performed to achieve a goal. When a goal is identified, this provides the skeleton for understanding and recalling what is required to accomplish it. Tasks are represented as a hierarchy of goals in a way similar to GOMS, but this approach takes the concept further by distinguishing between different forms of task knowledge. As a sibling method to GOMS this approach is considered not suitable for the same reasons.

A search of the literature using the criteria above resulted in zero matches, indicating this approach has not been used in a dental context at the time of writing. More broadly, it is claimed that the approach has been used a number of times but these examples were carried out for private clients meaning that the details of its use is not shared in the literature (Jonassen et al., 1999, p. 197).

B.2.13 Other approaches to task analysis

The following approaches were also investigated, but were considered inappropriate for use in this project as they either do not provide a structured task analysis approach or are more suited to interrogating and structuring existing sources of information:

- Syntactic Analysis
- Conceptual Graph Analysis
- Master Chart Method
- Matrix Analysis

- Repertory grid technique
- Fault Tree Analysis
- Business Process Modelling

All approaches were searched for using Web Of Science for any evidence that they had been applied in a dental context. All queries returned zero results with the exceptions of Matrix Analysis, Repertory Grid Technique and Fault Tree Analysis. Repertory Grid Technique and Fault Tree Analysis returned 17 and 3 results respectively, however, manually reviewing the abstracts of the results showed that these were not applied for task analysis in a dental context. Matrix Analysis returned 94 results but, this is a broader term used for example in *euclidean distance matrix analysis*, *extracellular-matrix analysis* and *co-occurrence matrix analysis*. When this term was AND'd with *task analysis* zero results were returned.

Appendix C

Task Analysis Session Schedules

This appendix describes the schedules for the Individual Simulation Sessions and Group Consensus Meeting. Copies of the participant data collection sheet and task decomposition form used during the Individual Simulation Session are also included.

C.1 Individual Session Schedule

Consent

- Offer a copy of participant information sheet and allow time to read/re-read if needed
- Draw attention to the recording equipment and section on video recording in the participant information sheet (including when the recordings will be deleted)
- Ask if any questions
- Ask participant to sign consent form

Session structure

- Record participant background information on participant data sheet
- Start recording the session

Read to participant:

In this simulation session, after presenting the scenario background, I will ask you to describe the main steps involved in preparing a tooth for restoration. You will then be asked to carry out the procedure in line with the steps you list, describing out loud what you are doing and thinking whilst working. At the end of each step there are a few questions – some you may cover whilst you are describing the step but it will give us an opportunity to make sure that nothing of interest has been missed.

Scenario Background

Read to participant:

An intra-oral examination, confirmed by appropriate radiographs, has identified that a patient requires the removal of caries and subsequent restoration of a lower molar. In consultation with the patient, it has been agreed that an amalgam restoration will be placed.

Thinking about how you would carry out this procedure, what do you consider to be the main (3 to 6) tasks or goals in order to prepare that tooth (caries removal and tooth preparation) up to the point of (but not including) the restoration?

- Write down each task in sequential order on a Post-It note – one per task and attach them to the edge of the simulator.

Starting with the first task – would you say that this can be broken down into any sub-steps or goals? If so, please list them.

Think-aloud

- Ask participant to seat themselves at the simulator

Read to participant:

The simulator has an exercise featuring a carious tooth. I would like you to prepare that tooth exactly as you would if this were a real case - one step at a time using the main tasks you have just described.

As you carry out a step, can you try and verbalise what you are doing and what you are thinking whilst working. For example, where you are paying attention, what you are looking out for, which tools you are using and why, what you are thinking about generally and so on.

Assume that the listener has zero dental knowledge but is keen to understand; provide as much detail as possible. Once you have completed a step, let me know and then we'll pause to talk about it in more detail if needed.

You are free to make any adjustments to the simulator as you normally would, but please think-aloud whilst doing so to capture the reasons behind why you are doing so. Finally, if at any point whilst using the simulator you experience any discomfort, please let me know and we can pause or end the session.

Closing

Read to participant:

Thank you for your time and your help with this project. Before we finish, is there anything that you consider important, or that a student should know in order to carry out this procedure that hasn't come up during this session?

I will now write up what you have described today and share it with you for your own use. You aren't expected to, but if you want to provide any corrections they would be appreciated.

C.2 Participant Data Form

Participant ID:
To be completed by a member of the research team

GDC Registrant Type:
Including speciality(ies) if applicable

Number of years since first registration:
To the nearest year, how long has it been since you qualified?

Number of years teaching experience:
How many years have you been involved in teaching?

C.3 Task Decomposition Form

Participant ______

Task/Step:

Goals:

"Describe the goals of this step"

Equipment and Motions:

"What equipment did you use and why? "How did you use these - describe the motions"

Attentional Focus:

"Where was your attention focused whilst carrying out this step"

Decisions and Cues:

"What decisions did you use make and what cues did you rely upon to make these decisions whilst carrying out this step (including that the step was complete)

Background Knowledge:

"In order to complete this step, what background knowledge did you rely upon?"

Common Mistakes:

"What are common mistakes novices make during this step?", "What are the consequences of these mistakes?"

Attributes of performance:

"What are the identifiable criteria or attributes of a successful execution of this step?"

VR Simulation Differences:

"What are the main differences between carrying out this step on the simulator just now when compared to the same procedure in the real world?"

"Were any of these differences significant and how did you have to accommodate them?"

Difference	Accommodation and significance

C.4 Group Consensus Meeting Schedule

Prior to meeting

- Compile draft group-consensus document and estimate required duration of focus group.
- Arrange convenient time for meeting between all participants and share with a private calendar event using everyone's University Google Calendar.
- Email participants, noting that whilst an individual may recognise their own contributions during the discussions, all contributions have been anonymised so no other participants will know the source of any given description. Remind participants that they are free to withdraw from the study if they do not wish to continue

Focus Group

Welcome, ground rules and consent

- After all members have joined the online meeting

Read to participants:

Thank you to everyone for joining this meeting. As noted in the Participant Information Sheet and the consent form, this session will be recorded and a copy of the recording kept until the conclusion of this project. Unless anyone has any objections I will start the recording now and proceed with the formal welcome and ground rules.

- Start recording

Read to participants:

We are now recording. Thank you for agreeing to participate in this group discussion. I have taken each of the individual descriptions produced from the task you carried out using the simulators and attempted to merge them into a first draft of a consensus between you all. I will first, present this draft in its current form pointing out areas where I was not able to simply merge your descriptions, then we will go through each of these differences, plus any other other interesting aspects, to create a group consensus description of the task. In the draft consensus, you may recognise aspects of your individual description but unless you reveal this fact, no other participant will be aware *that you are the source - it may even be the case that you were not the only person to describe the task in that way.*

Clearly, given the nature of this discussion your comments cannot be anonymised from each other, but I would ask each of you to:

- *Respect each others' opinions*
- *Avoid interrupting or talking over each other*
- *Maintain confidentiality - we will be following the Chatham House Rule so whilst you may use information from the discussion (for example you may find the other perspectives useful in your own teaching practice) you must not reveal the source.*

Please could you all indicate that you agree to these ground rules by nodding your head or giving a thumbs up.

- Wait for participants to indicate agreement

I will now present the draft group consensus, after which we will discuss each of the individual differences and come to a consensus description of the task.

Focus Group Discussion

- The lead investigator will present the draft group-consensus task description
- After the overview, for each difference noted the lead investigator will:
  - Present the description from the differing perspectives
  - Ask the participants to discuss and agree on a description for the step. This could be achieved by, selecting one of the existing individual descriptions, combining aspects of multiple descriptions or the creation of a new description.
  - After a consensus has been reached for the step, the decision will be noted and the next difference discussed.
  - If a unanimous description cannot be reached for a given step, participants will be informed all descriptions will be documented and the step annotated to state that a consensus could not be reached.
- After all differences have been discussed, the lead investigator will summarise the group consensus task description and ask for confirmation that the consensus task description was accurately captured.
- Finally, participants will be asked if they feel that there is anything missing from the consensus task description and be permitted a final discussion if necessary.

Conclusion of Meeting

Read to participants:

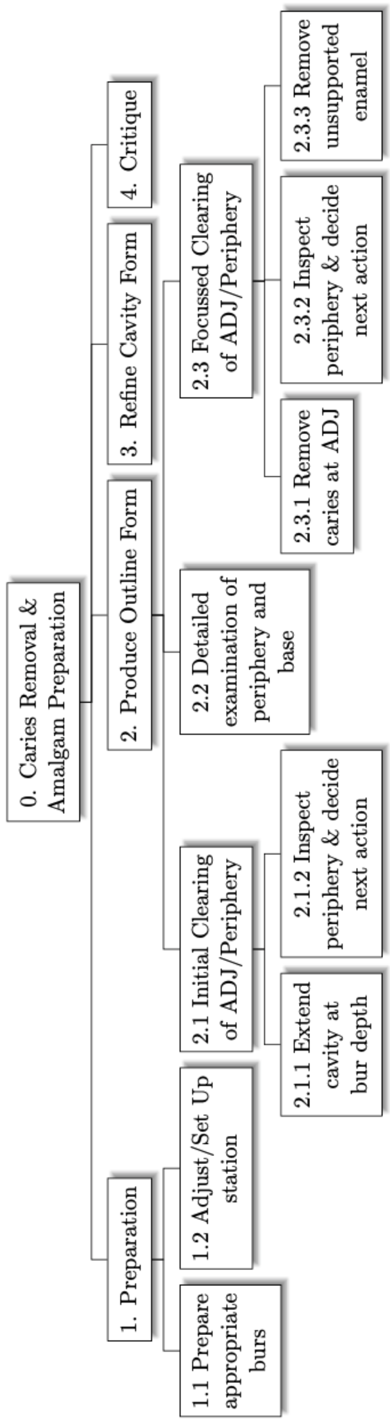
Thank you for your participation in this focus group. This approach and the results from your contributions form part of a wider project and will be written up and discussed in the context of that. However, if you would like to receive a draft copy of the group consensus task description that we have just produced for your personal use, please let me know and I will provide a copy as soon as I am able.

Appendix D

Task Description of Caries Removal and Cavity Preparation for an Amalgam Restoration

This appendix provides the full Task Descriptions produced from the individual participants and the combined Consensus Task Description as agreed by all participants.

D.1 Individual Descriptions - Participant 1



Participant 1, Individual Task Description

0 : Caries Removal and Amalgam Preparation

Task Plans/Operations

Plan 0: 1 – 2 – 3 – 4 – EXIT

1 : Preparation

Task Plans/Operations

Plan 1: 1.1 – 1.2 in any order

1.1 : Prepare appropriate burs

Task Plans/Operations

- Select a high speed handpiece with a pear shaped bur

Notes

This stage is significantly different in the VR context, so the differences are presented first, and then the attributes of performance/common mistakes are provided for the real world context.

VR Simulation Differences

- A major difference between the simulation and the real world is that choosing the burs that will be used during the procedure is a more definite step in the real world. The simulator provides a full menu of options with immediate access and the cost of switching between them approaches zero. In the real world, there are a number of factors that are present:
  - The burs required would need to be planned in advance so that they can be retrieved from the dispensary before starting. This introduces an experiential aspect whereby the burs that are most likely to be needed for a procedure are known.

- There is a cost associated with using a bur (either because they are single-use and disposable or the costs to sterilise it after use). If many burs are used to perform the procedure that will increase the cost of providing that treatment.
- Switching between different burs and different handpieces incurs a time cost in the real world, so an efficient operator will work with the smallest possible subset of burs to achieve the desired result.
- Swapping between burs requires handling them. Doing so repeatedly results in an increased risk of cross-infection.

It is desirable for a student to become proficient at knowing which burs they want. If they are employed at a very well stocked private clinic, they may be presented with lots of choices. Equally, in a less well stocked situation, knowing a good 2nd choice if their 1st choice is not available is desirable. If a student develops a good idea which burs they want, they will not waste time deciding between them.

- The menu of burs presented on the simulator is of an indeterminate scale. A real bur catalogue would print the burs at 1:1 or show a scale bar so that the size of the real instrument is obvious. In the simulator, burs that look bigger on the menu can be smaller once selected. This makes selecting a bur more difficult than it should be.

Attributes of performance

- A high speed handpiece and a pear-shaped bur should be selected - this shape bur will achieve most of the design parameters required so should be selected in most cases:
  - It automatically provides undercut and rounded internal lines. The undercut is needed to provide mechanical retention for the restoration so a bur that naturally affords this is desirable. The rounded internal lines avoid stress concentrations which can cause the tooth to crack.
  - The prepared cavity for a well retained amalgam restoration should not be too shallow (or it will be weak). The shoulder of the bur is approximately 2-3mm so this provides a guide to the appropriate depth. The position of the shoulder against the surface of the tooth can be used as a depth gauge cutting at roughly the average depth of the enamel.
- A rosehead bur should be selected for use in a low speed handpiece during later stages. A rosehead bur allows for more gentle removal of caries with a reduced risk of drilling too deeply and exposing the pulp.
- An efficient operator could carry out the procedure with as few as the two burs listed above. A larger cavity could indicate that two rose head burs (a large and

a small) would be useful, but the most efficient selection of burs is the smallest number that can produce the desired result with the least amount of switching between them.

Common Mistakes

- Selecting too many sizes and designs of burs and constantly swapping between them. This can be driven by an uncertainty of what the student is wanting to achieve or simply a perception that if lots of burs are available then lots are useful.
- Use too small of a bur, resulting in more passes being required to remove the material. This produces a poorer finish and will lead to a bigger cavity than would have been created if the correct size was used.

1.2 : Adjust/Set up station

Task Plans/Operations

- Come into contact with the simulated tooth with the bur (without turning the handpiece on)
- Attempt to get a secure finger rest on the simulator's finger rest stage
- Adjust height of simulator and chair to give an ergonomic working position
- If necessary, calibrate the simulator

Notes

None

VR Simulation Differences

- Difficult to get a realistic finger rest
- The design on the simulator introduces a tendency to rest the side of the hand on top of the simulator. This would not be possible in the real world.

2 : Produce Outline Form

Task Plans/Operations

Plan 2: 2.1 – 2.2 – 2.3 – EXIT

2.1 : Initial Clearing of ADJ/Periphery

Task Plans/Operations

Plan 2.1: 2.1.1 – 2.1.2; Loop until test at 2.1.2 is satisfied

2.1.1 : Extend cavity at bur depth

Task Plans/Operations

- During the 1st iteration of this step
  - Introduce bur into the centre of the lesion to the depth of the shoulder of the bur
  - Maintaining a constant depth, move laterally and antero-posteriorly to open the cavity up to the predicted initial width of the preparation
  - After removing the desired tissue, stop and inspect the periphery of the prepared cavity: 2.1.2
- During the 2nd to nth iteration
  - Apply a gentle pressure with the bur and brush laterally along the ADJ.
    - This is a gentle brushing motion to sweep away the caries, not a cutting motion to create a hole.
  - Focus should be on the bur and the line of brown caries at the ADJ as it disappears/is removed. The goal is to remove as little tissue as possible.
  - Observe the taper of the cavity that is being created, if it does not have a retentive form, tilt the bur to restore the undercut. This allows for the taper to be re-introduced whilst the cavity is being extended rather than adding it as a discrete step afterwards. If excessive undercut is introduced and needs to be removed, a upwards sweeping motion (rather than lateral) with the bur can remove it.
  - After extending the cavity a small amount, inspect the periphery and decide next action: 2.1.2

Notes

A school of thought is that the operator should switch to a rosehead bur in a low speed handpiece during the nth iteration of this loop. However, whilst there is a lot of enamel it is less efficient to do so. The rosehead would instantly create unsupported enamel which would need to be removed. Careful use of the high speed will remove both the caries and enamel until the line of caries at the ADJ is beginning to dissipate.

Attributes of performance

- Maintaining a rest and making as small movements as possible to carefully remove the tissue
- Maintain the orientation of the bur as perpendicular to the occlusal surface of the tooth
- Whilst a smooth outline should be created, the shape of the cavity is not important at this stage. The caries determines the shape, so successful performance is achieved by closely following the pattern of the caries as presented
- The bur should be held at a constant depth, growing the cavity wider as the ADJ caries is removed without removing deeper tissue.

Common Mistakes

- Students expect there to be more movement when preparing the tooth so struggle to make the small motions
- When clearing the ADJ, learners often remove too much tissue, creating too big a cavity well beyond how far the caries had spread laterally
- Switch to the rose-head too early and create undercut that needs to be removed
- Give too much attention to the periphery and lose track of the depth, creating either a deeper cavity than needed or partially withdrawing, creating a cavity which is too shallow.
- Lack of confidence leading to the student working slowly and running out of time in the clinical session
- Too much confidence and working too quickly, causing the student to remove too much tissue, shortening the life of the tooth.

VR Simulation Differences

- It is possible to drill from any angle in the VR context, meaning that the tooth can be accessed from angles that would not be possible in the real world. It is arguable if this should be possible in a teaching environment.
- Constant focus on the depth of the diamonds/shoulder of the bur is required - it is very difficult to judge the depth of the bur in VR. Whilst this cue is used in the real world, it must be relied upon more here because the other cues to judge depth such as 3d vision or the shadowing within the cavity are not present.
- The handpiece is simply on/off in the VR context. It does not have a variable speed control as would be present in the real world. To avoid over-cutting the pedal must be "pulsed" to mediate the speed of tissue removal.
- The simulator tends to cut rapidly unexpectedly so this encourages the user to be a bit more gentle.
- Because there is no restriction on the rotation of the view, it is more natural to use the camera position to view the preparation. This allows views of the preparation that would not be possible in the real world and discourages the use of the mirror.
- The simulator allows a learner to reset and try the exercise again. This is not possible in the real world so a student should be acutely aware that what they have cut, they have cut and there are no 2nd chances.

2.1.2 : Inspect periphery & decide next action

Task Plans/Operations

- Zoom in to give a closer view of the preparation
- Rotate the view 360° to inspect the periphery
- If significant caries can be seen at the ADJ and there is still plenty of enamel on the occlusal surface, zoom back out to a realistic size then return to 2.1.1 and extend the cavity.
- If the ADJ is approaching the point where it is cleared and only small dots of caries are visible or the lateral spread of the caries is further than expected necessitating more careful management, carry out a detailed examination: 2.2

Notes

None

Attributes of performance

- The ADJ should be *only just* cleared. The presence of some superficial dots of carries at the periphery indicates that the removal is at this threshold.

VR Simulation Differences

- In the real world the mirror would be used to inspect the periphery. In the simulator, the use of the view rotation is preferred because it is easier.
- The ability to inspect around the preparation is much greater than can be achieved in the real world.
- There is no ability to wash/dry the preparation in VR as would be done in the real world
- It is necessary to use the bur to physically infer the shape of the preparation because the lack of shadowing and 3d means that it is difficult to visualise its shape without touching it.
- There is no way to distinguish the hardness of the caries at the ADJ. In the real world a probe or excavator would be used to determine if it is carious or simply brown/stained tissue as an input to the decision to extend the cavity.
- The single-colour representation of the caries and the lack of being able to probe the tissue means the decision to remove more tissue is partly determined by an estimate of spread based on experience rather than the use of the cues that would be present in the real world.

2.2 Detailed examination of periphery and base

Task Plans/Operations

- View all aspects of the preparation and judge if:
  - $\frac{2}{3}$ of the periphery are completely clear of caries, with just small dots remaining. A mental note that these will need to be returned to should be made.
  - The walls of the preparation that require further caries removal require this to be done using a more delicate instrument. If no, these can be extended by resuming at 2.1.1

  - Depth is even and at the minimum requirement for the type of restoration being placed.
    * NB: The presence of caries on the floor of the cavity is acceptable at this stage.

- If satisfied with the preparation so far, continue with 2.3.

2.3 : Focussed Clearing of ADJ/Periphery

Task Plans/Operations

Plan 2.3: 2.3.1 – 2.3.2 – 2.3.3; Loop until test at 2.3.2 is satisfied

2.3.1 : Remove caries at ADJ

Task Plans/Operations

- Select a low speed handpiece
- Select a small rosehead bur that will permit the removal of a small margin of caries below the ADJ.
- Identify the area from 2.2 where caries must be carefully removed
- Use the bur (without turning on the handpiece) to feel the area that will be cut; rehearsing the motion.
  - This will give an idea of what that motion will feel like so that it can be executed when vision is impaired once the water spray starts
- Using small movements, remove the caries at the ADJ.
  - The motion with the rosehead is more *applied* than *brushed* along the ADJ. If the bur sinks in or material is readily removed, this is a sign that more work is necessary. If the material being removed is more like dust, suggests that only affected/stained dentine remains.

Notes

None

VR Simulation Differences

- The simulator does not display water spray. In the real world, the water spray obscures the view of what is being cut. Because this is missing from the simulation, it gives an unrealistic idea of what can be seen whilst operating.
- The simulator models that the low speed handpiece is completely unable to remove enamel. In reality, the rosehead would chip away at the enamel rather than removing small amounts. When working close to the ADJ the low speed can be prevented from removing tissue that would be possible in the real world because enamel has become involved.
- The texture of the material being removed is not represented in the VR environment. Highly demineralised tissue is leathery or waxy as it is removed. This gradually transitions to a dusty texture as it approaches healthy dentine. An operator should be paying attention to the nature of the material being removed as a guide to when to stop removing.
- The colour of the caries does not change and is represented as a single colour. In the real world the brown line of caries would transition from a dark brown to a lighter brown to a chalky white. In chronic caries, the chalky/demineralised tissue at the periphery can have a long transition back to healthy tissue creating uncertainty as to when to stop extending. In the simulation, the decision is simply to stop when the brown caries has been removed. This does not prepare the learner for how to deal with this aspect in the real world.

2.3.2 : Inspect periphery & decide next action

Task Plans/Operations

- After removing the caries at the ADJ with the low speed:
  - Inspect the periphery to establish if enamel has become involved preventing the removal of caries at the ADJ.
  - Determine if there is a significant amount of unsupported enamel
- If either of the above tests are met, extend the access: 2.3.3

2.3.3 : Remove unsupported enamel

Task Plans/Operations

- Select a high speed handpiece with a pear shaped bur

- Rehearse the motion of removing the desired enamel
- Carefully brush the enamel away, focussing on the areas where caries is still present at the ADJ
- Once almost all caries at the ADJ has been removed, shift focus to refining the cavity and continue at 3.

Notes

None

Attributes of performance

- The ADJ will be almost completely clear and just small dots of caries remaining

VR Simulation Differences

- When approaching this stage of the caries removal, there is almost always some chalky tissue at the extent of the carious spread. This is not included in the simulation.

3 : Refine Cavity Form

This phase overlaps 2.2 and there is an organic transition between the two. The focussed clearing of the ADJ leads to a progressively closer approximation of the final cavity. Once the caries is cleared (or almost cleared) at the ADJ the focus shifts to achieving other design criteria for the cavity.

Task Plans/Operations

- Visually check all aspects of preparation for desirable attributes of final cavity form:
  - Ensure that there is no unsupported enamel
  - Ensure an approximately 90° cavo-surface angle
  - Ensure a retentive form and that the cavity is of a suitable depth (minimum of the cutting head of the pear shaped bur - approximately 2-3mm)

- No sharp edges are present at the margins
- ADJ is fully cleared

- Where any of the above criteria are not met, using the high speed handpiece and a pear shape bur, refine the cavity form to ensure all desirable attributes are present.
  - NB additional focus should be given to the angulation of the bur at this stage to ensure that the desirable attributes are produced.
- Evaluate the base of the cavity to ensure it is relatively smooth and flat. If it is not, use a low-speed handpiece to introduce these
- When the cavity has been refined, critique the performance: 4

Notes

None

Attributes of performance

- Careful refinement of the final form, overlapping the last iteration of the caries removal at the ADJ stage, allows for additional tooth tissue to be preserved. By planning ahead, the tissue removal required to satisfy one criteria can be done in such a way to satisfy others. If these were carried out as discrete steps it is possible that additional tissue would need to be removed to achieve the same result. For example as the final caries at the ADJ is removed, it can be done to create the necessary retentive undercut.
- The lack of gradient in the colour or density of the caries means that a desirable attribute (in VR only) is to leave some caries in the base of the tissue to illustrate that consideration has been given to management of affected dentine. It indicates a high level of handpiece control (it is much easier to remove everything, but a thin layer of caries at the base requires a degree of skill). It would be a substantial improvement to the simulation experience if affected tissue was included.

VR Simulation Differences

- In the real world an excavator would be used to evaluate the density of caries at the base of the cavity. This is not present in the simulation so some carious tissue is left as a notional representation of the affected dentine.

- Because the caries does not vary in colour, this leads to behaviour that would not occur in the real world. The presence of a dark carious mark at the very periphery of the preparation indicates that it is demineralised tissue that should be removed, however, in the real world it is likely that this tissue would be lighter/white in colour (if visible at all). This mismatch of expectations encourages the user to remove the dark-coloured caries even though that coloured material would never be there in the real world.
- The lack of depth cues (3d vision, shadows etc.) tend to make the user cut deeper than they would in the real world.

4 : Critique

Task Plans/Operations

- Review the cavity and apply the attributes of performance to self-critique the preparation
- Identify aspects of the performance that went well or badly
- Reflect on any contributing factors that may have led to these aspects occuring. Consider what would be done differently (or should be repeated) next time.
- Decide if this performance is clinically acceptable?
  - If not, can it be made more clinically acceptable or is it the case that you should have stopped sooner?

Notes

This is an important step; without recognising aspects of good or bad performance, it is difficult for the learner to improve.

Attributes of performance

- ADJ has been cleared to no more than 0.5mm beyond the spread of the caries
- Cavity is at a depth of at least 2mm and has an approximately flat and smooth floor
- Cavity has a smooth periphery

- Cavity is as small as possible, laterally and depth of preparation
- All unsupported enamel has been removed
- The cavity retentive

Common Mistakes

- A student should know the criteria, but being able to cut them and recognise them is more challenging

VR Simulation Differences

- The simulator reveals extra information (such as the cross-sections) that would not be available in the real world, this allows additional opportunities to critique performance

D.2 Individual Descriptions - Participant 2

Participant 2, Individual Task Description

0 : Caries Removal and Amalgam Preparation

Task Plans/Operations

Plan 0: 1 – 2 – 3 – 4 – 5 – 6 – EXIT

Notes

None

Common Mistakes The following are common mistakes that relate to the procedure as a whole rather than to an individual step:

- Making incorrect decisions on if the tooth should be operated on at all
- Preparing a cavity that is too deep or too wide when the pre-operative radiograph indicated that it should be small
- Lack of preparedness for carrying out the procedure

VR Simulation Differences Broad simulator differences that apply throughout:

- No water spray
- No aspirator
- No rubber dam
- No soft tissue management e.g. risk of cut tongue
- No patient factors to be managed
  - Fidgety in the chair
  - Nervous
  - Worries about feeling the procedure or getting water down their throat
  - Needing to go to the toilet
  - etc.
- Tactile feel isn't quite right, the enamel isn't hard enough and there isn't any change/difference in the density/feel of the dentine as you progress through the infected tissue.

1 : Preparation

Task Plans/Operations

Plan 1: 1.1 – 1.2 – EXIT

1.1 : Assessment

Task Plans/Operations

- Review other restorations & consider patient treatment history
- Review radiograph to get an idea of the likely size of the cavity/restoration
- Administer local anaesthetic
- Isolate tooth undergoing treatment with rubber dam

Notes

None

Attributes of performance

- Appropriate assessment has been performed to guide the basic decision of if the tooth should be operated on at all - e.g. has the caries spread into the dentine or would it be more appropriate to monitor the lesion and/or treat it with fluoride.

VR Simulation Differences

- The assessment step is not properly represented the simulation experience, but in the real world a thorough assessment is vital. Knowledge of a patient's caries history can be a useful guide to predict how the treatment is likely to progress: If no other restorations are present, it is likely that this will be a superficial restoration; if all the 6's are already restored it's likely that it will be more substantial. Additionally, previous encounters with the patient can be instructive - is this patient compliant, will they accept rubber dam etc. These factors help inform and prepare the operator for what to expect so that appropriate preemptive measures can be taken.

- Radiographs are not available yet these are vital information for planning the procedure and informing what to expect.

1.2 : Set up simulator

Task Plans/Operations

- Zoom in to a size approximating actual size of the natural tooth
- Rotate view orientation to place the operator at approximately '10-to' on a clock face. This is the approximate the position the operator would be seated when accessing a lower-left molar as featured in the scenario
- Establish a secure finger rest

Notes

None

Attributes of performance

- The presence of a secure finger rest to support the handpiece whilst in use is crucial for accurate movements and connects the operator with the patient. Failure to have a secure finger rest whilst using the handpieces would be considered unacceptable

2 : Access to Lesion

Task Plans/Operations

Plan 2: 2.1 – 2.2 – EXIT

2.1 : Select appropriate bur

Task Plans/Operations

- Select a small pear shaped diamond bur of preferred size and a high speed handpiece

Notes

The specific size and design of small bur used will be determined based on an individual operator's preferences based on experience.

Attributes of performance

- Appropriate design small diamond grit bur and a high speed handpiece is selected in the simulator

2.2 : Access the cavity through enamel

Task Plans/Operations

- Introduce the bur into the cavity aligned along the long axis of the tooth
- Entry point should be at the most obvious point on the occlusal surface. This will be the area that is already cavitated, dark/black, stained or was identified from the radiograph.
- "Paint" away the carious enamel, extending laterally to open it up and provide access
  - The enamel is only 2-3mm deep so the bur will drop into the softer decayed dentine below. The depth at this stage should be controlled to limit the depth to approximately the depth of the cutting head of the bur.
- Once the enamel has been broken through and the access has been extended sufficiently to provide access for the next step then stop.

Notes

None

Attributes of performance

- The access cavity is just big enough to access the caries to permit the use of a rosehead bur during the next step

- Access should be made at the the correct point on the surface of the tooth that will give access to most of the caries
- Bur should be oriented along the long axis of the tooth and not angulated when cutting.

Common Mistakes

- Access in wrong place, just accessing at the middle of the tooth rather than where the main body of the caries is located
- Create too big an access cavity. NB Lateral extension is performed to provide access only, it should not extend to remove all visible caries, some of this may be staining rather than decay requiring removal.
- Novices sometimes believe they are removing caries with the high speed handpiece at this step (should be using the rosehead steel bur)
- Poor/uncontrolled access with the high speed can lead to iatrogenic pulpal exposure so students should be very careful when using this handpiece.

3 : Manage Caries at ADJ

Task Plans/Operations

Plan 3: 3.1 – 3.2 – 3.3 – 3.4; Loop until test at 3.3 is satisfied then immediately EXIT

3.1 Select appropriate bur

Task Plans/Operations

- Select a low speed handpiece
- Select the largest steel rosehead available for the task
  - In the first instance, this will be the largest bur that will fit through the access cavity cut in the preceding step
  - On subsequent passes after extending the access, the size selected may reduce in line with the tapering of the amount of caries towards the limit of its spread below the enamel. A larger bur will stop readily removing decayed tissue as it is also in contact with healthy tissue, if soft carious dentine is still present

at the ADJ a smaller size rosehead should be selected in order to fully clear the ADJ.

3.2 : Remove accessible caries at ADJ

Task Plans/Operations

- In a circular shape motion (without going deeper), remove caries present under the ADJ, adjacent to the enamel, by "dragging" the rosehead around the walls of the tooth.
- When accessible caries has been removed, proceed to 3.3.

Notes

None

Attributes of performance

- The focus of this step is to remove caries at the ADJ only, so removing material from the base of the cavity is considered undesirable at this stage.
- The use of a larger bur is much safer and less likely to penetrate deeper into the tooth so would be considered desirable
- At the limits of the spread of the caries the dentine becomes more creamy in colour, this is a visual indicator of when this step is complete so further removal should be done with caution as it may indicate too much tissue is being removed.

Common Mistakes

- Often students use too small a bur for this step. Small burs are not efficient for removing caries and can risk drilling a hole deeper into the tooth. A larger bur will only remove softer dentine with a much-reduced risk of cutting deeper.
- Using high speed handpiece to remove caries and enamel (both at once) resulting in a cavity wider than would be created using the correct technique

3.3 : Evaluate ADJ

Task Plans/Operations

- Assess all walls of the cavity to see if any soft caries is still present and extending beneath the enamel at the ADJ
  - Soft caries should be removable by the low speed handpiece. When the tissue being removed is seen to change to a dry "dentine dust" rather than sticky black/brown decayed tissue, that is an indicator that the area is clear
- If caries is still present beneath the enamel but cannot be accessed, continue to 3.4, if the ADJ is completely clear, EXIT and continue with 4

Notes

None

Attributes of performance

- ADJ must be clear and free of caries in all cases. Failure to identify that caries is still present would be considered poor performance of this step.
- Healthy dentine is often visually lighter and more *creamy* coloured once clear of caries, however it can be stained. The stained tissue is similar in density to healthy dentine and cannot be removed with the low speed handpiece. This stained tissue should be retained.

Common Mistakes

- Fail to identify that the ADJ has not been cleared and progress to clearing floor of cavity prematurely

3.4 : Extend Access

Task Plans/Operations

- Select a high speed handpiece

- Select a diamond grit bur
- Carefully remove enamel by painting it away using a brushing motion on the wall of the tooth where caries is still present until the caries at the ADJ can be accessed.

Notes

NB the cavo-surface angle is not considered at this stage, the extension is only to increase the available access to caries at the ADJ, not to refine the cavity design.

Attributes of performance

- The minimum amount of enamel has been removed so that access to the ADJ can be regained.
- Only the fissures with infected tissue are removed, extending the preparation through all fissure patterns is an outdated approach

Common Mistakes

- Removing too much enamel with the high speed handpiece and attempting to remove caries at the same time as extending the access. This risks leading to a cavity bigger than necessary.

4 : Manage Caries at Pulpal Floor

Task Plans/Operations

- Refer to radiograph to assess likely depth of caries/risk of pulpal exposure
- Decide if using a total or partial caries removal approach. This will be guided by the depth of spread shown in the radiograph
- Select a large steel rosehead in a low speed handpiece
- In a circular motion remove the soft caries from the base of the cavity. The same tactile/visual cues as when clearing the ADJ can be used when deciding how much material should be removed.

Notes

None

Attributes of performance

- Only soft decayed tissue at the pulpal floor is removed, healthy or stained-only tissue should be retained.
- A large bur is selected to reduce the risk of drilling too deeply.
- Tissue is removed gradually, bur is not rapidly "plunged in":
  - If the caries is deep into the tooth, it may be possible to perform partial caries removal and negate the need for a RCT. However, if the bur has been plunged in too quickly resulting in a pulpal exposure, this option is no longer available.

Common Mistakes

- Remove all tissue without discrimination even if extent of decay is seen to be close to the pulp on the preoperative radiograph
- Not knowing when to stop removing tissue - missing visual and tactile colour/consistency cues that indicate a transition to material that can be retained:
  - Infected tissue has:
    - Wet or soggy consistency
    - Dark black/brown, sometimes bluey colour
  - Affected tissue is:
    - Drier
    - Lighter in colour.
- Failure to refer to the radiograph to gain insight into how close the decay has spread towards the pulp so that the patient can be properly warned of the risks.
- Producing a cavity that is too deep or too wide when pre-op radiograph indicated it would be a small preparation

VR Simulation Differences

- No transition in colour between infected and affected dentine
- No transition in density between infected and affected dentine

5 : Refine Cavo-Surface Angle

Task Plans/Operations

- Select a high speed handpiece and appropriate diamond grit bur
- Assess all aspects around the margin of the cavity
  - Any enamel which is not supported by dentine should be removed
  - Where cavo-surface angle is not 90 degrees, carefully remove enamel until a 90 degree angle is achieved.

Notes

None

Attributes of performance

- Angle at cavo-surface on all aspects should be 90 degrees
- No unsupported enamel has been left

Common Mistakes

- Failure to carry out this step at all
- Leaving large overhangs.
  - These will be weak and have an increased risk of fracture. Also, the presence of large overhanging tissue may indicate that the ADJ has not been fully cleared. This can result in a restoration that is already compromised when it is brand new.

6 : Final Review

Task Plans/Operations

- Dry the cavity
- Look closely at all aspects of the preparation and evaluate performance in terms of the attributes below
- Using the probe:
  - Feel for any unsupported areas of enamel, indicating that further enamel should be removed
  - Feel for any soft dentine indicating that caries has not been fully managed
- If acceptable, EXIT. If not acceptable, carry out steps above to correct any unsatisfactorily completed steps as identified.

Notes

The following attributes of performance are what a tutor would be looking for when evaluating the procedure as a whole:

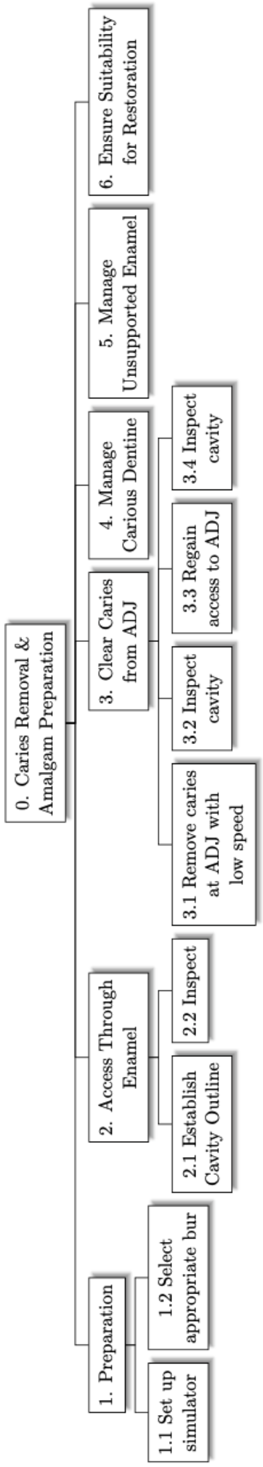
Attributes of performance

- In reference to the pre-operative radiograph, the preparation is conservative - the amount of tissue removed is not excessive.
- Is the cavity caries free or has it been managed correctly.
  - If a total caries removal approach is adopted, has all caries been removed?
  - If using a partial-caries removal approach, has infected tissue been misclassified as affected and requires removal.
- Is the prepared cavity targeted, did it follow the caries - or is it a stylised shape e.g. perfectly flat floor.
- Has the cavo-surface angle been correctly prepared and is it appropriate for the material being used to restore the tooth.

VR Simulation Differences

- No ability to dry the cavity (or any suggestion of wet/dry tissue at all)
- Probe does not accurately model interactions with caries, e.g. does not model soft/hard tissue interactions

D.3 Individual Descriptions - Participant 3



Participant 3, Individual Task Description

0 : Caries Removal and Amalgam Preparation

Task Plans/Operations

Plan 0.1 - **Default:** 1 – 2 – 3 – 4 – 5 – 6 – EXIT Plan 0.2 - **Small Cavity:** 1 – 2 – 3 – 4 – 5: Optional – 6 – EXIT In a small lesion, 5 may be completed as a consequence of 3.1 so is optional Plan 0.3 - **Grossly unsupported:** 1 – 2 – 3 – 5 – if additional caries revealed Do 3 else Do 4 – 5 – 6 – EXIT;

1 : Preparation

Task Plans/Operation

Plan 1: 1.1 – 1.2 – EXIT

Notes

Assume starting with the task itself, all pre-op and rubber dam etc. already done and starting with the task itself

1.1 : Set up simulator

Task Plans/Operations

- Hide gloved hand, it is easier to visualise the task without it
- Zoom in to a size approximating viewing the tooth using 2x (ballpark) loupes
- Adjust height of simulator to ensure an ergonomic working position
- Ensure that the handpiece is held correctly and that secure finger rest can be achieved, calibrating the simulator if necessary.

Notes

None

Attributes of performance

- Simulator is at a height where good posture can be achieved
- A secure finger rest has been established

1.2 : Select appropriate bur

Task Plans/Operations

- Select a 2mm pear shaped diamond bur in a high speed handpiece

Notes

The selection of a 2mm pear shape is a personal preference of the operator. Ideal when preparing for amalgam but also suitable for composite. However, 2mm is the average depth of the enamel and a pear shape bur affords a natural undercut so (when used correctly) will naturally produce desirable attributes in the prepared cavity.

2 : Access Through Enamel

Task Plans/Operations

Access through enamel (to establish cavity outline and provide access for further caries removal):

Plan 2: 2.1 – 2.2, if test at 2.2 satisfactory EXIT, else Do 2.1 to make necessary adjustments.

2.1 : Establish Cavity Outline

Task Plans/Operations

- Identify the carious tooth
- Examine the carious lesion visually, and with reference to the radiograph, decide on the access point

- Starting at one end (personal preference, starting in centre also reasonable) of the visible carious enamel, introduce the bur to a depth of 2mm and follow the outline in a single movement, removing all visible carious enamel
  - Focus should be on the brown/discoloured area on the surface of the tooth and to maintain a constant depth of 2mm with an angulation of the bur along the long axis of the tooth

Notes

None

Attributes of performance

- Cavity is generally smooth around the margin
- Is generally of an even depth (but soft caries may make this difficult).
  - At this stage the carious lesion is not being followed into the depth of the cavity (although the outline in enamel is being followed), simply that access to it is being achieved.
- Able to see and feel the ADJ just below the enamel.

Common Mistakes

- Make the cavity too big at this stage. In an occlusal lesion (as in this case), the radiograph will not routinely show the depth that the caries extends to. Therefore it is essential to follow the outline, by removing the carious enamel exclusively. Further extension will be completed when clearing the ADJ. (as covered in 3)
- Use inappropriate angulation. Students often struggle to angle the bur to the correct plane of the tooth.
- Struggle with depth and knowing how deep 2mm is. Should use cues from bur if it is a 2mm cutting depth. Alternatively can take measurements from the bur or use a perio probe.

VR Simulation Differences

- Perio probe is not available in the simulator (measurements can be made in the cross-sectional views, but this would not be available in the real world.)

2.2 : Inspect

Task Plans/Operations

- Inspect the access cavity ensuring that can see:
  - The ADJ has been uncovered
  - The underlying dentine
  - The carious enamel has been removed
- If the inspection is satisfactory, continue to 3, if not return to 2.1 and make any necessary corrections

3 : Clear Caries from ADJ

Task Plans/Operations

Gain insight into size of lesion and establish sound caries free margin:

Plan 3: 3.1 – 3.2; if test at 3.2 shows that ADJ is clear EXIT, else Do 3.3 – 3.4; if test at 3.4 shows caries at ADJ Do 3.1, else EXIT

Notes

Students find this step difficult to visualise and often find it confusing, it takes a while to know what is required

3.1 : Remove caries at ADJ with low speed

Task Plans/Operations

- Select a low speed handpiece and a small rosehead bur.
  - A small rosehead is selected initially as the initial cavity is small. The purpose is to create a small caries free margin at the ADJ, not to remove gross caries at the base at this stage.
- Starting with the buccal wall, introduce the cutting head into the cavity and using a brushing motion, remove accessible carious tissue along the wall of the cavity at the ADJ.

  - To remove the caries with the low speed handpiece, more lateral force is applied than was used with the high speed in the previous step
  - The bur will not be at the full depth of the cavity, just deep enough to remove the caries at the periphery at and immediately below the ADJ. The aim is to create a caries-free margin. The cavity should not be made deeper at this stage.
  - Caries below the ADJ is accessible if it can be removed below the enamel without tilting the bur away from its alignment along the long axis of the tooth

- Mentally, divide the tooth up into the 4 internal walls and ensure that this process has been carried out on all sides.
- When all of the accessible caries has been removed, proceed to 3.2

Notes

None

Attributes of performance

- Only lateral caries at the ADJ has been removed (cavity has not been made deeper - just wider)
- Accessible caries at ADJ has been removed or a caries free margin has been established
- Correct angulation has been maintained

Common Mistakes

- Use too big a bur and remove too much tissue
- Don't remove enough tissue to clear ADJ
- Incorrect angulation of the bur or inserting it too deep below the enamel, removing too much deeper carious (and often sound) tissue - pushing out/diverging at the base of the cavity.

VR Simulation Differences

- The simulation of carious tissue is not quite accurate. The simulator in principle allows adequate demonstration of removal of caries at the ADJ. One problem is that there is a fixed boundary between enamel and dentine - which of course is anatomically correct. However, it is often difficult to visualise whether the remaining caries is within the enamel or dentine on the simulator. The instruments required for each substrate are different (enamel - high speed, dentine - slow speed) - this is often the same in real life (very minimal stains in enamel can be removed with a slow handpiece), however removing very small pieces of enamel on the simulator is risky as it is much easier for the bur to slip. Removing small amounts of enamel in real life is a safe and easy to perform procedure and visualising the difference between enamel and dentine is also easier.

3.2 : Inspect cavity

Task Plans/Operations

- Inspect each of the 4 walls of the cavity to establish:
  - If caries can be seen extending beyond the accessible areas of the ADJ, or
  - If visibility of the ADJ is impaired due to unsupported enamel
- If either of the above conditions are met, Do 3.3. If no more caries is present at the ADJ and there is no impairment to visibility of the ADJ, EXIT.

Notes

As caries is removed under the enamel, there will start to be overhanging/unsupported material. This overhanging tissue prevents complete visibility of the ADJ and prevents access to further carious tissue.

Common Mistakes

- Failure to identify that caries is still present at the ADJ (not checking all aspects of the cavity) through incomplete removal of grossly unsupported enamel

3.3 : Regain access to ADJ

Task Plans/Operations

- Select a high speed handpiece and a pear shaped diamond bur
- For each of the walls identified in 3.2: Using a careful brushing motion, not pressing on, and open up the cavity removing the unsupported enamel until ADJ is visible again
  - It is helpful to rehearse the exact motion that will be made with the high speed handpiece without turning it on or coming into contact with the tooth

Notes

None

VR Simulation Differences

- The lack of shadows can make it difficult to see unsupported enamel
- The high speed handpiece can sometimes slip in a way that would not occur in the real world

3.4 : Inspect cavity

Task Plans/Operations

- Inspect the newly accessible areas of the ADJ to see if caries is present
- If satisfied that only healthy/sound dentine of a yellow colour can be seen at the ADJ, Do 4. If caries is still seen, Do 3.1 else EXIT

Notes

None

Attributes of performance

- A small margin of caries-free dentine can be seen all the way around the cavity at the ADJ

4 : Manage Carious Dentine

Task Plans/Operations

- Select the largest rosehead bur that will fit into the cavity.
  - A large bur is less likely to perforate deeper into the soft carious tissue
- Using a brushing motion across the base of the cavity "peel away" the carious tissue from the pulpal floor.
  - Very little downward pressure should be applied so that a smooth floor can be established.
- In determining the depth of the cavity, distinguish between caries-infected dentine and caries-affected dentine on the pulpal floor:
  - Infected dentine appears wet, soft and can be scooped out. This should be removed.
  - Affected dentine has been affected by the carious process (demineralisation/loss of minerals) but is harder/drier. This tissue can be left in place, following partial caries removal approach (it can be remineralised/repaired).

Notes

None

Attributes of performance

- Important to ensure all previous steps are complete, if not it is good practice to return to complete prior to beginning this stage of carious removal.
- Base of the cavity is smooth

VR Simulation Differences

- The simulation does not have variable densities/textures/colours of caries. This prevents this aspect of the caries removal process from featuring in the simulation.
- Currently, the teacher asks students to remove an estimated portion of this caries and simply leave some caries (brown tissue on the simulator) deep in the base of the cavity in place, assuming that it is likely to be affected.

5 : Manage Unsupported Enamel

Task Plans/Operations

- Select a high speed handpiece A pear shaped diamond bur is routinely used by the operator for moderate-large amounts of unsupported enamel.
- Very fine burs (rugby or needle burs) can be selected for small amounts of unsupported enamel.
- Using a careful brushing motion, not pressing on and addressing each wall of the cavity in turn:
  - Remove any enamel that is not supported by dentine
  - Ensure all enamel edges are smooth and rounded
  - NB as above, the motion can be rehearsed to increase the accuracy when removing the material.

Notes

None

Attributes of performance

- Any jagged, rough or sharp points are prone to fracture, these should be smoothed off
- All enamel is supported by dentine

Common Mistakes

- Applying too much force and removing too much tissue. Should be a gentle painting motion to carefully remove the intended tissue.
- Failure to identify unsupported enamel. A probe can be run against the wall from base of the dentine up to the enamel, the presence of unsupported/overhanging enamel will be indicated by the probe catching on the overhanging tissue.

6 : Ensure Suitability for Restoration

Task Plans/Operations

- For amalgam, check that the cavity has the following attributes:
  - Relatively flat floor
  - Presence of undercut to retain the material
  - No rough/sharp points or areas of unsupported enamel (this should have already been addressed during 5)
- If these are not present, make any necessary adjustments

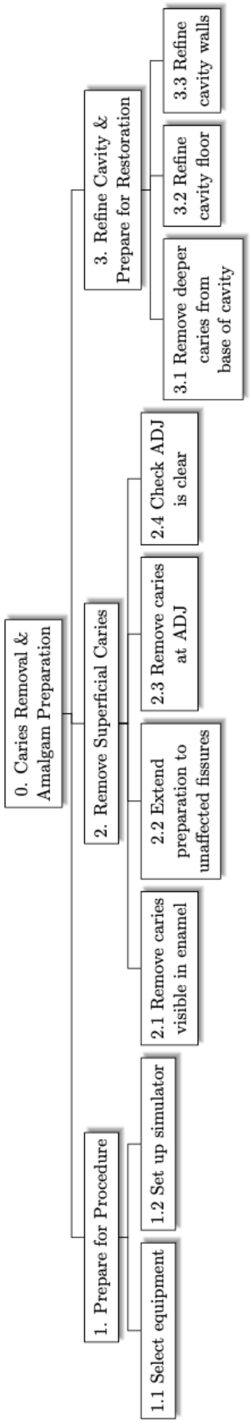
Notes

None

Attributes of performance

- The restoration has been checked and the required attributes for the restorative material have been confirmed to be present.

D.4 Individual Descriptions - Participant 4



Participant 4, Individual Task Description

0 : Caries Removal and Amalgam Preparation

Task Plans/Operations

Plan 0: 1 – 2 – 3 – EXIT

Notes

During the real procedure, prior to the goal "Prepare for Procedure", the procedure would be explained to the patient and informed consent obtained.

1 : Prepare for procedure

Task Plans/Operations

Plan 1: 1.1 – 1.2 – EXIT

Notes

None

VR Simulation Differences During preparation for the real procedure, the tooth requiring treatment would be identified and verified via a radiograph and local anaesthetic administered. The scenario as presented and the current simulator experience skip some of these steps. It would be a useful addition to the simulation experience if an X-Ray of each exercise were available.

1.1 : Select equipment

Task Plans/Operations

- Choose the correct handpiece, bur type and size for use in the procedure.

Notes

Both high speed and low speed handpieces are used in the caries removal and cavity preparation, the pre-selected pear-shape bur in a high speed handpiece was considered acceptable to start the procedure but a low speed would be preferred closer to the pulp.

There is an element of personal preference in the selection of burs. These preferences are seen to develop in senior year students, but for early stages, students should be recommended which burs to use.

Clinically it is possible to use a single bur to carry out the entire procedure. The main aim is to get the desired result, and a skilled operator can achieve this with a limited selection of burs. However, a poor choice of bur would reduce the likelihood of achieving the desired result (particularly for novices) so burs and handpieces appropriate for the task should be preferred.

VR Simulation Differences Often caries removal would include the use of hand instruments. These are not available in the simulation so do not feature in the detailed task description.

1.2 : Set up simulator

Task Plans/Operations

- Zoom view to high magnification
- Attempt to achieve a comfortable finger rest and seating position
- If a comfortable operating position can be can be achieved, EXIT
- If a comfortable operating position cannot be achieved, adjust the chair to desired height and/or calibrate the handpiece on the simulator so that a secure rest can be achieved on the finger rest stage.

Notes

None

2 : Remove superficial caries

Task Plans/Operations

Plan 2: 2.1 – 2.2 – 2.3 – 2.4; If test at 2.4 satisfactory EXIT else Do 2.3.

Notes

The focus for this phase is on the superficial caries at and around the depth of the ADJ.

VR Simulation Differences The bur in the simulator has a tendency to not cut and then jump, rapidly removing undesired material. A slip like this would only happen in the real world if a patient moved unexpectedly, but would be mitigated by the presence of a secure finger rest.

Suspected cause of this behaviour is that the shank of the bur is in contact with the tooth and holding the cutting head away from the surface of the tooth. This is hard to see and the simulation does not provide any tactile feedback that this is happening, so the operator perceives it as not providing sufficient pressure to cut through the material. In response, they increase the amount of force applied and in doing so, slightly change the angulation enough for the virtual cutting head to come into contact with the surface of the tooth. Once in contact the bur begins cutting again and removes a large amount of material before the operator is able to regulate the excess force being applied.

Common Mistakes

- Students often fail to remove all the caries that should be removed
- Students often focus too much on the base of the cavity at the expense of clearing ADJ.

2.1 : Remove caries visible in enamel

Task Plans/Operations

- Introduce a pear shaped diamond bur in a high speed handpiece into the tooth at the centre of the visible caries to the depth of the bur's cutting head.

- Using a mesial-distal sweeping motion at an even depth, remove all visible caries-infected enamel, paying attention to the depth of the bur to ensure that it does not cut deeper than intended.

Notes

None

Attributes of performance

- Need to be as conservative as possible and only remove the superficial visible layer of caries. Removing deeper caries at this stage is undesirable.

2.2 : Extend preparation to unaffected fissures

Task Plans/Operations

- At the depth of the cutting head (approximately 2mm) of the bur, follow the fissure pattern of the tooth, extending the cavity to remove tissue from all unaffected fissures.

Notes

This is performed because amalgam is not a conservative material and this must be done for retention of the placed restoration.

Attributes of performance

- The extended cavity is at an even depth with a smooth floor.
- The walls of the extended cavity are convergent to retain the restoration.

2.3 : Remove caries at ADJ

Task Plans/Operations

- Taking each internal wall of the tooth in turn:

- Using the side of the bur at a depth equal to the cutting head, open the cavity laterally, removing the caries present at the ADJ in a sweeping motion
- Once satisfied that the caries has been removed at the ADJ, attend to the next internal wall.

Notes

Main focus at this stage is paying attention to ADJ and ensuring that it has been thoroughly cleared.

Common Mistakes

- Not having a stable finger rest whilst working

Attributes of performance

- No caries should have been left where the enamel and dentine meet at the ADJ
- Size of cavity is determined by how far the caries has spread, so a cavity that closely matches the extent of the spread is desirable.
- Some caries can be left in the base of the cavity at this stage.

2.4 : Check that ADJ is clear

Task Plans/Operations

- Ensure that there is no caries present at the ADJ on any internal wall of the tooth.
- If satisfied, move on to task 3, if not repeat 2.3

Notes

None

Attributes of performance

- A thorough inspection should reveal that there is no caries at the ADJ. The presence of caries after inspection would be considered a poor performance of this step.

3 : Refine cavity and prepare for restoration

Task Plans/Operations

Plan 3: 3.1 – 3.2 – 3.3 – EXIT

Attributes of performance

- To create the desired cavity, only the carious tissue and the minimum amount of healthy tissue should be removed. Excessive tissue removal would be considered a poor performance.

3.1 : Remove deeper caries from base of cavity

Task Plans/Operations

- Using a large rosehead in a low speed handpiece remove the softer deeper caries using a distal-mesial stroking motion across the floor of the cavity.

Notes

None

VR Simulation Differences In the real world the caries in the base of the cavity would be tested/removed with a hand instrument. If the caries is "hard" or "scratchy" this can be considered "affected" tissue and left in place.This allows the procedure to be as conservative as possible. As the simulator does not have hand tools, a low speed was used for illustrative purposes in order to complete the task.

Attributes of performance

- All infected caries has been removed
- Affected caries on the floor of the cavity has been left in place
- The base of the cavity should be relatively smooth, but this can be tidied up in step 3.2.

3.2 : Refine cavity floor

Task Plans/Operations

- Using a large rose head bur in a slow handpiece, remove any uneven elevations in the base of the cavity.

Notes

The aim is to even-out the floor of the cavity without making it any deeper. It is a trade off - some deeper marks must be left in place because removing the tissue to even out the floor around them would make the cavity deeper and risk exposing the pulp.

3.3 : Refine cavity walls

Task Plans/Operations

- Using a fissure bur, remove/smooth out any unevenness from the walls of the cavity
- If not already present, introduce a convergent angulation to the walls

Notes

Ideally, the walls should be cut to be convergent whilst removing the caries, but if this is not achieved, a fissure bur can be used to remove tissue to introduce this design feature.

VR Simulation Differences

- A fissure bur would remove enamel in the real world if additional pressure is applied, resulting in a continuous angle across the dentine and enamel. The simulator does not allow any enamel removal when a slow speed handpiece is selected so this had to be accommodated by selecting an alternative bur in a high speed. However, the participant did feel it is useful to keep this distinction to remind the student that the high speed should be used for enamel removal.
- Overall, struggled to get a finish to the same quality of finish as is possible in the real world

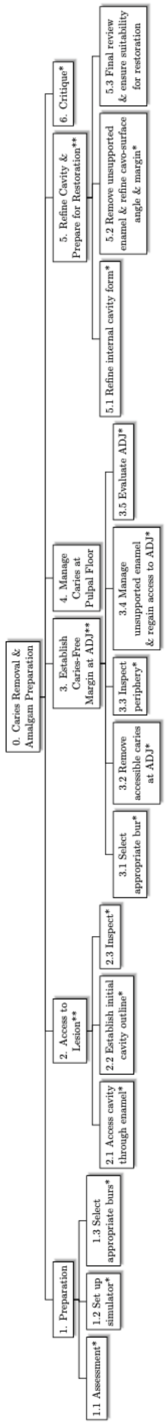
Attributes of performance

- Walls should be smooth
- Walls should be convergent
- No further material has been removed from the floor of the cavity. Only material from the walls is removed at this stage.
- Walls should have a minimum width of 2mm so that they are strong enough to not fracture. Walls thinner than this would indicate that a different class of restoration is required.

Common Mistakes

- Spend too much refining the shape leading to a bigger/wider cavity than was needed
- Using the wrong handpiece/bur to refine the cavity. For example, using a rosehead to refine the cavity walls can lead to dimples in the walls and undermine the enamel, the fisher bur "finds" the walls better and avoids this from happening.

D.5 Consensus Description



Consensus Task Description

0 : Caries Removal and Amalgam Preparation

Task Plans/Operations

Plan 0: 1 – 2 – 3 – 4 – 5 – 6 – EXIT

Notes

None

VR Simulation Differences Broad simulator differences which apply throughout in the VR environment:

- No water-cooling spray from the handpiece
- No aspirator
- No rubber dam
- No soft tissue management e.g. risk of cutting the patient's tongue
- No patient factors to be managed such as:
  - Explaining the procedure and obtaining consent
  - Patient Nervousness
  - Compensating for patient moving or being "fidgety" during the procedure
  - Patient needing to take a break e.g. to go to the toilet
  - etc.
- Differences in tactile sensations; the tissues do not fully represent how they feel in the real world and lack any variance in density or colour e.g as caries-infected dentine transitions through infected to affected to healthy dentine or a white chalky outline in the enamel.
- Finger rest is difficult to establish and does not resemble what would be used in the real world, the hardware design also promotes using the side of the hand on top of the simulator which would not be possible with a real patient.

1 : Preparation

Task Plans/Operations

Plan 1: 1.1 – 1.2 – 1.3 – EXIT

1.1 : Assessment

Task Plans/Operations

- Identify tooth requiring treatment and perform visual inspection to identify presence and location of caries
- Review other restorations & consider patient treatment history to suggest how far the caries is likely to have spread. Previous experience with the patient will also be a factor for how to approach the treatment e.g. compliance, accept rubber dam etc.
- Review radiograph of tooth requiring treatment to estimate and visualise the size and extent of the needed cavity/restoration. Based on the depth of the caries, decide if there is a risk of pulpal exposure.

Notes None

Attributes of performance

- The tooth requiring treatment has been correctly identified
- The visual assessment and an examination of the radiograph have appropriately confirmed that the caries has spread into the dentine and requires intervention.

Common Mistakes

- Making an incorrect decision to operate on the tooth when it should have been monitored and/or treated with fluoride instead.

VR Simulation differences

- The assessment step is not fully represented the simulation experience. In the real world a thorough assessment is vital as knowledge of a patient's caries history can be indicative of how the treatment is likely to progress, for example if this is the patient's first restoration it is likely that it will be a superficial; if all the 6's are already restored it might suggest poor oral hygiene and that a substantial restoration will be required. Additionally, previous encounters with the patient can be instructive - is this patient compliant, will they accept rubber dam etc. These factors help inform and prepare the operator for what to expect so that appropriate preemptive measures can be taken. If conducting an assessment were an overt part of the simulated exercise it would add a great deal of richness to the experience.
- Exercises in the simulator do not provide a radiograph. These are vital information for planning the procedure and informing what to expect. The cross sectional views give some of this information but in a way that is dissimilar to the real world, missing out on an opportunity to develop this skill.

1.2 : Set-up simulator

Task Plans/Operations

- Zoom in to a magnification of approximately the what would be seen when using 2x loupes or to approximately the actual size of the real tooth.
- Rotate view orientation to place the operator at approximately '10-to' on a clock face. to approximate the view the operator would see when seated to access a lower-left molar.
- Holding the handpiece in a modified pen grip, attempt to get a secure finger rest on the simulator's finger rest stage and an ergonomic working position. If both can be achieved, EXIT
- If a comfortable operating position cannot be achieved, adjust the height of the simulator and/or chair
- If a secure finger rest cannot be established, calibrate the simulator

Notes None

Attributes of performance

- The correct height of the simulator and a properly adjusted chair permits an ergonomic posture to be achieved whilst it is in use. Failure to correctly set these will contribute to musculo-skeletal health problems long-term. Focus on this aspect will be to ensure 120° angles at the major joints. Paying attention to this angle at the elbow will assist in setting the height of the simulator. NB This cannot be assessed by the simulator.
- A secure finger rest is essential to support the handpiece whilst it is in use and enables accurate movements to be made. It also connects the operator to the patient in case of any unexpected movements. The lack of a secure finger rest whilst using the handpiece would be considered unacceptable. The focus of establishing a secure finger rest is that the tooth in the exercise can be accessed when the ring and little finger are in stable contact with the finger rest stage in the simulation. NB the presence/absence of a finger rest cannot be assessed by the simulator.

Common Mistakes

- Failure to adjust the chair or height of the simulator to achieve a correct seating position, simply reusing the position left by the previous user.
- Not recognising that a secure finger rest is not present or resting the side of the hand on the top of the simulator.
- Failure to calibrate the simulator hand position when necessary

VR Simulation differences

- It is difficult to get a realistic finger rest
- The design of the simulator introduces a tendency to rest the side of the hand on top of the simulator. This would not be possible in the real world.
- Calibration (and recognising problems with calibration) are not real-world considerations
- The simulator allows levels of zoom that far exceed what would be used during most restorations. This can give the student an unrealistic idea of what they will be able to see when carrying out the procedure in the real world. A degree of operator preference is acceptable for the exact level of zoom, but this should be limited to 2.5x whilst working on normal caries removal exercises. The higher levels of zoom should only be used when inspecting the performance afterwards.

- The ability to enable/disable the gloved hand divided opinion. Some tutors saw value in the presence of the hand because it represented the common difficulty of gaining good visibility around the operator's hands and instruments. However, others saw it as simply a distraction that gets in the way, removing the student's focus from learning the process of caries removal.
- The visibility available in the simulation means that the mirror is not necessary. Whilst the real-world visibility of the tooth featured in the exercise is often adequate without a mirror, the mirror would be used for retraction and inspecting the mesial aspect. Not requiring the mirror for either of these purposes means that the student misses out on practice at bi-manual coordination and again provides a misleading view of the ability to directly inspect all aspects of the preparation. However, it could equally be claimed that this is simply part of the learning process and granting easier access promotes scaffolding of skills over time.

1.3 : Select appropriate burs

The simulation and how this would occur in a clinical situation differ here. In the simulation context, the user should follow the operations as stated, but in the real world this step would involve selecting and retrieving the selection of burs that are most likely to be required to carry out the procedure from the dispensary.

Task Plans/Operations

- From the handpiece menu, select a high speed handpiece
- From the high speed burs menu select a small/2mm pear shaped diamond bur.

Notes None

Attributes of performance

- Whilst the selection of a 2mm pear shaped bur is a personal preference expressed by the participants, the shape of this bur will assist in achieving many of the design parameters of the final preparation:
  - It automatically provides undercut and rounded internal lines. The undercut is needed to provide mechanical retention for an amalgam restoration so a bur that naturally affords this is desirable. The rounded internal lines avoid stress concentrations which can cause the tooth to crack.

  - The prepared cavity for a well retained amalgam restoration should not be too shallow (or it will be weak). The shoulder of the bur is approximately 2mm so the position of the shoulder against the surface of the tooth can be used as a depth gauge corresponding with the approximate average depth of the enamel.

Therefore, whilst selecting the 2mm pear shape is not mandatory, it is indicative of an awareness of the target form of the preparation.

Common Mistakes

- Selecting a bur that is too small for use in the next step. A bur which is too small will require more passes to be made in order to remove the material. This produces a poorer finish and will lead to a bigger cavity than if the correct size was used.

VR Simulation differences

- A major difference between the simulation and the real world is that choosing the burs that will be used during the procedure is a more definite step in the real world. The simulator provides a full menu of options with immediate access and the cost of switching between them approaches zero. In the real world, there are a number of factors that are present:
  - The burs required would need to be planned in advance so that they can be retrieved from the dispensary before starting. This introduces an experiential aspect whereby the burs that are most likely to be needed for a procedure are known.
  - There is a cost associated with using a bur (either because they are single-use and disposable or the costs to sterilise it after use). If many burs are used to perform the procedure that will increase the cost of providing that treatment.
  - Switching between different burs and different handpieces incurs a time cost in the real world, so an efficient operator will work with the smallest possible subset of burs to achieve the desired result.
  - Swapping between burs requires handling them. Doing so repeatedly results in an increased risk of cross-infection.

It is desirable for a student to become proficient at knowing which burs they want. If they are working at a very well stocked private clinic, they may be presented with lots of choices. Equally, in a less well stocked situation, knowing a good 2nd choice if their 1st choice is not available is desirable. If a student develops a good idea which burs they want to achieve a given outcome, they will not waste time deciding between them.

- An efficient operator could carry out the procedure with as few as a single pear-shaped bur for use in the high speed handpiece and a rose-head bur for use in the low speed handpiece. A larger cavity may indicate that two rose head burs (a large and a small) would be useful, but the most efficient selection of burs is the smallest number that can produce the desired result with the least amount of switching between them. Failing to select appropriate burs (corresponding with the skill level of the operator) will reduce the likelihood of achieving the desired result.
- Often caries removal would include the use of hand instruments. These would be collected at this stage for use in the procedure but are not available in the simulation.
- It is difficult to estimate the size of the burs on the menus in the VR simulation. In the real world it is immediately obvious based on their actual size, but the menu in the simulator can be misleading where a bur that appears small in the menu is larger once selected.

2 : Access to Lesion

Task Plans/Operations

Plan 2: 2.1 – 2.2 – 2.3 – If test at 2.3 is satisfactory: EXIT else: Do 2.2

2.1 : Access cavity through enamel

Task Plans/Operations

- Examine the carious lesion visually, and with reference to the radiograph, decide on the access point. This will be the area that is already cavitated, dark/black stained or was identified as the area with the most decay from the radiograph.
- With reference to the available information, mentally estimate the extent of the spread of caries that is likely to be encountered.
- Move the handpiece towards the selected entry point on the occlusal surface. Angulate the bur along the long axis of the tooth, start the handpiece and make a controlled entry into the cavity at the selected entry point to the depth of the cutting head of the bur.

  - The enamel is only 2-3mm deep so the bur will drop into the softer decayed dentine below. Attention should be paid to the cutting head of the bur and the depth limited to this point on the bur.

- "Paint" away the carious enamel, with the side of the bur, extending laterally to open the entry point up and provide minimal access
- Once the enamel has been broken through and the access has been extended sufficiently for the next step: EXIT.

Notes None

Attributes of performance

- Access should be made at the point on the surface of the tooth that will give access to most of the caries
- Bur should be oriented along the long axis of the tooth and not angulated when cutting.
- Depth should be controlled to the depth of the cutting head of the bur
- Access should be small at this stage to provide access/visibility to the adjacent tissue

Common Mistakes

- Access in wrong place, just accessing at the middle of the tooth rather than where the main body of the caries is located
- Create too big an access cavity. NB Lateral extension is performed to provide access/visibility at the area with the most caries only, it should not extend to remove all visible darker tissue at this stage, some may be staining rather than caries requiring removal.
- Poor/uncontrolled access with the high speed can lead to iatrogenic pulpal exposure so students should be very careful when using this handpiece.

VR Simulation differences None

2.2 : Establish initial cavity outline

Task Plans/Operations

- Continuing from the access created above, using a gentle brushing or sweeping action with the side of the bur, follow the outline of the visible carious enamel; removing all caries in the enamel that has progressed through to the dentine. Staining on the surface of the tooth that does not progress through to the dentine should not be removed.
  - NB This is a gentle brushing motion to sweep away the caries, not a cutting motion to create a hole.
- The bur should be maintained at a constant depth at the shoulder of the bur (approximately 2mm) orientated along the long axis of the tooth. This depth will ensure that the enamel is cleared and access given to the underlying ADJ and dentine. Maintaining a secure finger rest and making as small movements as possible are important whilst removing the tissue.
- Whilst considering the predicted extent of the spread from 2.1, focus should be on removing the brown/discoloured tissue in the enamel and evenly maintaining the intended depth with correct angulation along the long axis of the tooth.
- When the visible carious enamel has been removed: EXIT

Notes None

Attributes of performance

- As skill develops, the initial access and establishment of the initial cavity outline will become a single operation without any break between the steps.
- Able to see and feel the ADJ just below the enamel all around the cavity.
- Cavity is generally smooth around the margin, however, the shape is not important at this stage. The caries determines the shape, so successful performance is achieved by closely following the pattern of the caries as presented
- Is generally of an even depth (but soft caries may make this difficult).
- Must be as conservative as possible at this stage, only removing the superficial visible caries - removing any deeper caries at this stage is undesirable. It is sufficient that access has been achieved.

- The angulation of the bur along the long axis of the tooth has been maintained throughout.

Common Mistakes

- Make the cavity too big at this stage. The only ‘known’ is the caries in the enamel, so this should be followed exclusively. Further extension will be carried out in later steps.
- Use inappropriate angulation. Students often struggle to angle the bur to the correct plane of the tooth.
- Struggle with depth. Students sometimes struggle to estimate how deep 2mm is. This can be assisted with cues from bur (if it is a 2mm cutting depth) or by taking measurements from the bur or a perio probe.
- Losing track of depth. Students often give too much attention to the periphery and lose track of the cutting depth. This leads to either a deeper cavity than needed or an unretentive shallow cavity.
- Making too big movements. Students expect there to be more movement in preparing a tooth so struggle to make the small motions required.

VR Simulation differences

- Perio probe is not available in the simulator so cannot be used to reinforce its role as a measurement tool (measurements can be made in the cross-sectional views, but this would not be available in the real world.)
- It is possible to drill from any angle in the VR context, meaning that the tooth can be accessed from angles that would not be possible in the real world. It is arguable if this should be possible in a teaching environment.
- Very difficult to judge the depth of the bur in VR so constant focus on the depth of the diamonds/shoulder of the bur is required. Whilst this cue is used in the real world, it must be relied upon more extensively here because the other cues to judge depth such as 3d vision or the shadowing within the cavity are not present.
- The handpiece is simply on/off in the VR context. It does not have a variable speed control as would be present in the real world. To avoid over-cutting the pedal must be “pulsed” to mediate the speed of tissue removal. This would not be done in the real world.

- The simulator can cut rapidly and unexpectedly so this encourages the user to be a bit more gentle.
- Because there is no restriction on the rotation of the view, it is more natural to use the camera position to view the preparation. This allows views of the preparation that would not be possible in the real world and discourages the use of the mirror.
- The simulator allows a learner to reset and try the exercise again. This is not possible in the real world so a student should be acutely aware that "what they have cut, they have cut" and there are no 2nd chances.

2.3 : Inspect

Task Plans/Operations

- Zoom in to give a closer view of the preparation
- Rotate the view 360° to inspect the periphery of the access cavity and confirm that:
  - The ADJ has been uncovered
  - The underlying dentine is visible
  - The carious enamel has been removed
- Identify where caries can be seen extending below the enamel at the ADJ
- If the inspection is satisfactory: Do 3, else Do 2.2

Notes None

Attributes of performance

- The initial cavity outline has been established and the above criteria have been confirmed.

VR Simulation differences

- In the real world the mirror would be used to inspect the periphery. In the simulator, the use of the view rotation is preferred because it is easier. Failure to reinforce the use of the mirror to inspect the periphery may be undesirable.

- The lack of shadows, homogenous colour of tissue-types etc means that the ability to inspect the preparation is much greater than can be achieved in the real world because the differences in colouration are more obvious.
- There is no ability to wash/dry the preparation in VR as would be done in the real world.
- It is necessary to use the bur to physically infer the shape of the preparation because the lack of shadowing and 3D means that it is difficult to visualise its shape without touching it.

3 : Establish Caries-Free Margin at ADJ

Task Plans/Operations

Plan 3: 3.1 – 3.2 – 3.3 : If no unsupported enamel preventing visibility or inaccessible caries detected at ADJ: EXIT else Do: 3.4 – 3.5 If satisfied: Do 3.1 else: EXIT.

As judgement and fine motor control develop with expertise, the number of iterations around these sub goals will reduce. An expert operator will be able to judge (drawing more reliably from the assessment and interpretation of the X-Ray) the extent of the carious spread. This enhanced skill and knowledge from a broader experience base allows them to correctly identify the enamel that must be removed to gain access to the caries below, completing in a single iteration what might require a novice many iterations to achieve. However, this is an efficiency that will develop with expertise and is not appropriate for novices to attempt during early skills development.

3.1 : Select appropriate bur

Task Plans/Operations

- Select a low speed handpiece from the handpieces menu
- Select an appropriately sized small rosehead bur, guided by:
  - The size of this will be limited by:
    - The current width of the opening
    - The depth of the access
    - The depth of the carious dentine visible on the cavity walls.

    Select the largest bur within these constraints that will not also remove caries from the base of the cavity at the same time.

  - A smaller size may need to be selected in subsequent iterations because the caries tapers towards the extent of the carious spread. The larger size will be in contact with healthy dentine which will prevent it from readily removing the decayed tissue; a smaller size will allow the ADJ to be fully cleared.

Notes None

Common Mistakes

- Select too big a bur which will remove too much tissue and simultaneously extend the depth prematurely
- Select too small a bur. Small burs are not efficient for removing caries and can risk drilling a hole deeper into the tooth. A larger sized bur will only remove softer dentine with a much-reduced risk of accidentally cutting a hole.

3.2 : Remove accessible caries at ADJ

Task Plans/Operations

- Mentally, divide the tooth up into the 4 internal walls then, starting with the buccal, address each wall in turn:
  - Identify the caries spreading at the ADJ that was noted during 2.3
  - Use the bur (without turning on the handpiece) to feel the area that will be cut; rehearsing the motion.
    * This will give an idea of what that motion will feel like so that it can be confidently executed when vision is impaired by the water spray
    * The bur will not be at the full depth of the cavity, just deep enough to remove the caries at the periphery at and immediately below the ADJ. The aim is to create a caries-free margin. The cavity should not be made deeper at this stage.
  - Using small movements, remove the accessible caries at the ADJ:
    * Caries at the ADJ is considered accessible if it can be removed below the enamel without tilting the bur away from its alignment along the long axis of the tooth.
    * To remove the caries with the low speed handpiece, more lateral force is applied than was used with the high speed in the previous step. The motion with the rosehead is more deliberately *applied* than *brushed* along the ADJ.

        * Attention should be given to the nature of the material being removed:
            · If the bur sinks in or material is readily removed, this is a sign that more work is necessary.
            · If the material being removed is more like dust, suggests that either only affected/stained dentine remains or sound dentine is interfering with the removal process. NB dentine can be stained but have a similar density to healthy dentine; this tissue should not be removed.
        * At the limits of the spread of the caries the dentine becomes more creamy in colour, this is a visual indicator of when this step is complete. Further removal should be done with caution as it may indicate too much dentine is being removed.
    - When all accessible caries has been removed, continue with the next wall until all 4 sides have been addressed.

- When all accessible caries at the ADJ has been removed from each of the 4 walls, EXIT

Notes None

Attributes of performance

- Only lateral caries at the ADJ has been removed (cavity has not been made deeper than it was at the end of Goal #2)
- Accessible caries at ADJ has been removed or a caries free margin has been established
- Correct angulation has been maintained and the base of the cavity has not been pushed out.

Common Mistakes

- Not having a stable finger rest whilst working
- Insufficient tissue is removed to clear ADJ to an acceptable margin
- Prematurely focussing on the base of the cavity at the expense of properly clearing the ADJ
- Incorrect angulation of the bur or inserting it too deep below the enamel, removing too much deeper carious (and often sound) tissue - pushing out/diverging at the base of the cavity.

- Using a high speed handpiece to remove caries and enamel (both at once) resulting in a cavity wider than would be created using the low and high speed together.
- Lack of confidence leading to the student working slowly and running out of time in the clinical session
- Too much confidence and working too quickly, causing the student to remove too much tissue, shortening the life of the tooth.

VR Simulation differences

- The colour of the caries does not change and is represented as a single colour. In the real world the brown line of caries would transition from a dark brown to a lighter brown to a chalky white. In chronic caries, the chalky/demineralised tissue at the periphery can have a long transition back to healthy tissue creating uncertainty as to when to stop extending. In the simulation, the decision is simply to stop when the brown caries has been removed. This does not prepare the learner for how to deal with this aspect in the real world.
- It is often difficult to visualise whether the remaining caries is within the enamel or dentine on the simulator. The instruments required for each substrate are different (enamel - high speed, dentine - slow speed) - this is often the same in real life (very minimal stains in enamel can be removed with a slow handpiece), however removing very small pieces of enamel on the simulator is risky as it is much easier for the bur to slip. Removing small amounts of enamel in real life is a safe and easy to perform procedure and visualising the difference between enamel and dentine is also easier.
- The simulator does not display water spray. In the real world, the water spray obscures the view of what is being cut. Because this is missing from the simulation, it gives an unrealistic idea of what can be seen whilst operating.
- The simulator models that the low speed handpiece is completely unable to remove enamel. In reality, the rosehead would chip away at the enamel rather than removing small amounts. When working close to the ADJ the low speed can be prevented from removing tissue that would be possible in the real world because enamel has become involved.
- The texture of the material being removed is not represented in the VR environment. Highly demineralised tissue is leathery or waxy as it is removed. This gradually transitions to a dusty texture as it approaches healthy dentine. An operator should be paying attention to the nature of the material being removed as a guide to when to stop removing.

- The bur in the simulator has a tendency to not cut and then jump, rapidly removing undesired material. A slip like this would only happen in the real world if a patient moved unexpectedly, but would be mitigated by the presence of a secure finger rest.

3.3 : Inspect periphery

Task Plans/Operations

- Carefully inspect the periphery across each of the 4 walls of the preparation in turn to assess:
  - If enamel has become involved preventing the removal of caries which can be seen at the ADJ
  - If visibility of the ADJ is impaired due to the presence of grossly unsupported enamel preventing verification that the ADJ has been fully cleared
  - Inspect the taper of the cavity that is being created, if it does not have a retentive form, attempt to re-introduce this whilst the cavity is being extended in Goal 3.4. This is more conservative than adding the retention as a discrete step afterwards and should be preferred where possible.
- If any caries can be seen extending beyond the accessible areas of the ADJ or visibility of the ADJ is impaired, Do 3.4 else, EXIT.

Notes None

Attributes of performance

- Main focus is paying attention to ADJ and ensuring that it is clear and free of caries. The goal here is to identify areas where enamel must be removed to restore access to the ADJ so that 1) further caries can be removed or 2) it can be verified as being cleared. Failure to identify unsupported enamel or that caries is still present would be considered poor performance of this step.
- The ADJ should be *only just* cleared. The presence of some superficial dots of carries at the periphery indicates that the removal is likely at this threshold.
- The presence of caries in the base of the cavity is not a concern at this stage
- Maintaining the correct retentive form during the procedure places an additional cognitive load on the novice. However, it is crucial throughout, therefore amalgam restorations are often introduced later in a clinical skills programme once other dependent skills of handpiece control and caries removal are more established.

Common Mistakes

- Failure to identify that caries is still present at the ADJ due to:
  - Not thoroughly checking all aspects of the cavity
  - Incomplete visibility/access to the ADJ due to the presence of grossly unsupported enamel
- Prematurely progressing to clearing the floor of the cavity

VR Simulation differences

- There is no way to distinguish the hardness of the caries at the ADJ. In the real world a probe or excavator would be used to determine if it is carious or brown/stained. This information would be used to inform a decision of if the cavity should be extended.
- The single-colour representation of the caries and the lack of being able to probe the tissue means the decision to remove more tissue is not fully based on the cues that would be present in the real world. Instead, removal is partly informed by an estimate of the likely extent of the spread and a judgement based on experience.
- The lack of shadows makes identification of overhanging enamel harder than it is in the real world.

3.4 : Manage unsupported enamel & regain access to ADJ

Task Plans/Operations

- Select a high speed handpiece and a pear shaped bur
- For each wall where access to the ADJ must be regained or unsupported enamel managed:
  - Rehearse the exact motion that will be required without turning on the handpiece or coming into contact with the tooth
  - Maintaining angulation along the long axis of the tooth and using a careful lateral brushing motion, without pressing on, paint away the enamel using the side of the bur using smooth strokes.
  - Continue to do so until the grossly unsupported enamel has been removed and ADJ is visible/accessible again on each of the walls.

Notes None

Attributes of performance

- The minimum amount of enamel has been removed so that access to the ADJ can be regained.
- The cavo-surface angle is not considered at this stage, the focus is to increase the available access to caries at the ADJ, not to refine the cavity design.
- Whilst not the primary focus of this stage, if the taper of the cavity was identified as not being retentive during Goal 3.3, the bur can be tilted slightly to restore the undercut whilst extending the cavity at the same time.

Common Mistakes

- Removing too much enamel with the high speed handpiece and attempting to remove caries at the same time as extending the access. This risks leading to a cavity bigger than necessary.

VR Simulation differences

- The lack of shadows can make it difficult to see unsupported enamel. This is much easier to identify in the real world.
- The lack of depth cues (3d vision, shadows etc.) tend to make the user cut deeper than they would in the real world.
- The high speed handpiece can sometimes slip in a way that would not occur in the real world

3.5 : Evaluate ADJ

Task Plans/Operations

- Visually inspect all aspects of the newly uncovered/accessible areas of the ADJ to see if caries is present and extending at the ADJ.
- If satisfied that only healthy/sound dentine can be seen at the ADJ Do 4; else Do 3.1

Notes None

Attributes of performance

- Size of cavity is determined by how far the caries has spread, so a cavity that closely matches the extent of the spread is desirable.
- A thorough inspection should reveal that a small margin of caries-free dentine can be seen all the way around the cavity at the ADJ
- Planning ahead is encouraged at this stage. If (based on just a small amount of caries remaining at the ADJ) the next iteration of the process will (probably) be the last, the form of the cavity can be refined whilst doing so. This permits future desirable attributes of the cavity being created at the same time as achieving the current goals, overall, preserving more tooth tissue than would be done as discrete steps. For example, as the final caries at the ADJ is removed, it can be done in such a way as to create the necessary retentive undercut.

Common Mistakes

- Failure to identify that caries has spread further at the ADJ

VR Simulation differences

- In the late iterations of this stage, there is almost always some chalky tissue at the extent of the carious spread. This is not included in the simulation so the learner is not prompted to look for this visual cue that they are reaching the limits of how far it has spread.

4 : Manage Caries at Pulpal Floor

Task Plans/Operations

- Refer to radiograph to re-assess likely depth of caries/risk of pulpal exposure
- Decide if using a total or partial caries removal approach. This will be guided by the depth of spread shown in the radiograph
- Select the low speed handpiece and the largest rosehead bur that will fit into the prepared cavity.

  - A large bur is less likely to unintentionally perforate deeper into the soft carious tissue

- Using a circular brushing motion across the whole base of the cavity “peel away” the soft carious tissue gradually from the pulpal floor.

  - Very little downward pressure should be applied so that any uneven elevations are removed and a smooth floor can be established.

- In determining the depth of the cavity, distinguish between caries-infected dentine and caries-affected dentine on the pulpal floor:

  - Infected dentine appears wet, soft and can be scooped out. This should be removed.
  - Affected dentine has been affected by the carious process (demineralisation/loss of minerals) but is harder/drier. This tissue can be left in place, following partial caries removal approach (it can be remineralised/repaired).
  - The same visual cues as when clearing the ADJ can be used in deciding which material is infected and should be removed.

Notes

None

Attributes of performance

- All soft infected caries has been removed from the pulpal floor
- Stained, caries affected dentine and healthy dentine has been retained
- Base of the cavity is smooth

  - Some deeper marks can be left in place where removing the tissue to even out the floor around them would make the cavity deeper and risk exposing the pulp

- Tissue is removed gradually in layers - the bur is not repeatedly “plunged in”. Where deeper caries is present it may be possible to perform partial caries removal and negate the need for an RCT. Plunging the bur in too quickly can result in a pulpal exposure, removing this opportunity.

Common Mistakes

- Remove all tissue without discrimination even if decay is seen to be close to the pulp on the pre-operative radiograph
- Not knowing when to stop removing tissue - missing visual and tactile colour/consistency cues that indicate a transition to tissue that is suitable to be retained
- Failure to refer to the radiograph to gain insight into how close the decay has spread towards the pulp so that the patient can be properly warned of the risks.

VR Simulation differences

- The simulation does not have variable densities/textures/colours of caries. These factors are used to inform decision making in the real world but cannot be used in the VR environment.
- Currently, the teacher asks students to remove an *estimated* portion of the caries and simply leave some behind in the base of the cavity in place as a token to represent the affected tissue. Leaving behind a token amount of caries demonstrates that the student is aware that this might be done in the real world and requires a higher degree of handpiece control. However, the decision is based on background knowledge rather than correctly encountering and recognising the cue in the simulation.
- In the real world the caries in the base of the cavity would be tested/removed with a hand instrument. If the caries is “hard” or “scratchy” this would be considered “affected” and left in place leading to a more conservative result. Because the simulator does not have hand tools and the density of the affected dentine cannot be tested, this part of the caries removal process cannot be experienced in the VR environment.
- A great deal of the checking and judgement made during the caries removal process is done via hand tools. The absence of an excavator and the fact that the probe does not differentiate between densities of caries means that the operator does not use the same process in VR as would be used when carrying out in real world. The simulation of the hand tools should be improved to permit a representation of these steps.
- Because the caries does not vary in colour, this leads to behaviour that would not occur in the real world. The presence of a dark carious mark at the very periphery of the preparation indicates that it is demineralised tissue that should be removed, however, in the real world it is likely that this tissue would be lighter/white in colour (if visible at all). This mismatch of expectations can persuade the user to

remove the dark-coloured caries even though that coloured material would never be there in the real world.

5 : Refine Cavity & Prepare for Restoration

Task Plans/Operations

Plan 4: 5.1 – 5.2 – 5.3 If satisfied at 5.3: EXIT, else: Do 5.1

Notes

This phase overlaps the previous phase and there is an organic transition between the two.

5.1 : Refine internal cavity form

Task Plans/Operations

- Check that a minimum width of 2mm of enamel is remaining around all walls of the cavity.
  - This minimum ensures that the walls are strong enough to not fracture. If less than this remains, it would indicate that a different class of restoration is now required.
- Check that the cavity is of a suitable depth (minimum of approximately 2-3mm) using an appropriately sized bur or a perio probe.
- Ensure that the base of the cavity is relatively smooth and flat.
- Visually check that the cavity has a retentive form with convergent angulation to the walls
- Check that the walls are smooth and do not have any unevenness
- If any of these criteria are not met, using the low speed handpiece and a large rosehead, make the appropriate modifications.
  - NB additional focus should be given to the angulation of the bur at this stage to ensure that the desired attributes are produced.

Notes None

Attributes of performance

- Walls should be smooth and convergent
- The cavity meets the minimum depth requirement
- The walls meet the minimum width requirement

Common Mistakes

- Over-refining the shape leading to a bigger/wider cavity than was needed

VR Simulation differences

- Overall, the cutting properties of the simulated tools mean that it is difficult to get a finish to the same quality of finish as is possible in the real world

5.2 : Remove unsupported enamel, refine cavo-surface angle & margin

Task Plans/Operations

- Test for the presence of unsupported enamel by using a probe:
  - Run it against the wall from base of the dentine up to the enamel, the presence of unsupported/overhanging enamel will be indicated by the probe catching on the overhanging tissue.
- If any unsupported enamel is identified:
  - Select a high speed handpiece and an appropriate bur:
    * A pear shaped diamond bur is often appropriate for moderate-large amounts of unsupported enamel.
    * Very fine burs (rugby or needle burs) can be selected for small amounts of unsupported enamel.
  - Using a careful lateral brushing motion, not pressing on and addressing each wall of the cavity in turn:

    * Remove any enamel that is not supported by dentine
    * Ensure all enamel edges are smooth and rounded
    * As above, the motion can be rehearsed to increase the accuracy when removing the material.

- Visually inspect all aspects of the margin for any rough edges of enamel or areas where the cavo-surface angle is not approximately 90°

- If any rough or sharp points are identified or the cavo-surface angle is not correct:

  - Carefully make the necessary corrections, removing enamel until a 90° angle is achieved or the rough edge is removed, using the same motion described for unsupported enamel above.

Notes None

Attributes of performance

- All unsupported enamel has been removed

- No rough or sharp points are present at the margin

- Angle at cavo-surface on all aspects is 90°

Common Mistakes

- Failure to carry out this step at all

- Failing to identify or leaving unsupported enamel:

  - This will be weak and have an increased risk of fracture.
  - The presence of large areas of overhanging enamel could indicate that the ADJ was not fully explored/cleared.

- Leaving jagged, rough or sharp points of enamel. These are prone to fracture, degrading the restoration

- Applying too much force and removing too much tissue. Should be a gentle painting motion to make the final refinements to the preparation.

VR Simulation differences None discussed

5.3 Final review & ensure suitability for restoration

Task Plans/Operations

- Dry the cavity
- Conduct a detailed review of the final preparation, looking closely at all aspects to ensure that the design criteria (for amalgam restorations) intended in the previous steps have been achieved:
  - An approximately 90° cavo-surface angle is present all around the margin
  - No sharp edges of enamel are present at the margins
  - ADJ is fully clear does not show any visual signs of caries
  - The floor is relatively flat
  - The preparation has a retentive form is of a suitable depth
  - Using the probe:
    * Feel for any unsupported areas of enamel, indicating that further enamel should be removed
    * Feel for any soft dentine indicating that caries has not been fully managed
- If the review is acceptable and all design criteria are met, EXIT; else: Do 5.1.

Notes None

Attributes of performance

- The cavity is caries free or has been managed appropriately:
  - If a total caries removal approach was adopted, all caries has been removed.
  - If using a partial-caries removal approach, infected tissue has been correctly identified and removed
- The preparation exhibits the required attributes for the selected restorative material listed above.

VR Simulation differences

- No ability to dry the cavity (or any suggestion of wet/dry tissue at all)
- The probe does not accurately represent the interaction with carious tissue because it does not penetrate the soft/lower density surface when pressed into it. This reduces the use of the probe in caries identification to a performative act in the VR environment.

6 : Critique

Task Plans/Operations

- Review the cavity and pre-operative radiograph and apply the attributes of performance below to self-critique the preparation
- Identify aspects of the performance that went well or badly
- Reflect on any contributing factors that may have led to these aspects occuring. Consider what would be done differently (or should be repeated) next time.
- Decide if this performance is clinically acceptable?
  - If not, can it be made more clinically acceptable or is it the case that you should have stopped sooner?

Notes

None

Attributes of performance

- In reference to the pre-operative radiograph, the preparation is conservative with a proportionate amount of tissue removal to that indicated on the radiograph.
- ADJ has been cleared to no more than 0.5mm beyond the spread of the caries
- The prepared cavity is targeted and followed the caries - it is not simply a stylised shape e.g. perfectly flat floor.
- Cavity is at a depth of at least 2mm and has an approximately flat and smooth floor

- Cavity has a smooth periphery
- All unsupported enamel has been removed
- The cavity retentive
- Cavity is as small as possible, laterally and depth of preparation. To meet the above criteria, the absolute minimum amount of healthy tissue has been removed.

Common Mistakes

- Failure to recognise the aspects of a good or bad performance. A student should know these criteria but being able to recognise them is more challenging. If they are not able to do so it will prove difficult for them to improve.

VR Simulation differences

- The simulator reveals extra information (such as the cross-sections) that would not be available in the real world, this allows additional opportunities to critique performance. In some contexts this is valuable, but can also provide a level of detail that exceeds what is clinically relevant.

Appendix E

Creating Simulator Exercises

This appendix details the process used to create each of the exercises listed in Chapter 7 using Blender v. 3.5.0 2023-03-29 (Blender Foundation, Amsterdam).

E.1 Creating the Exercise Models

Activity 1.1

To create Activity 1.1 (See Figure 7.1), a pre-existing tooth model segmented in to 3 separate .stl files representing the enamel, dentine and pulp was imported in to Blender. A *decimate* modifier was applied to each object with a ratio of 0.01 using collapse mode. This created the desired low-polygon aesthetic for the activities. To insert the ADJ text, first, a circle curve was added to the scene. This was positioned in the approximate location of the ADJ below the occlusal surface and scaled to an appropriate size. A text object with the repeating string "ADJ" was added and a *curve* modifier targeting the circle added to make the text follow the circle. The text was then scaled to prevent it overlapping when looping the circle. The text was then converted to a mesh and then subdivided a number of times via a loop cut tool to provide sufficient geometry for it to conform to the contours of the ADJ. The text object was then extruded from the centre point to reach the desired size.

To make the text conform to the contours at the ADJ a Grid was added and positioned just above the text object. A *surface deform* modifier with a strength of 0.7 was applied to the text object and targeted and bound to the grid. The two objects were then positioned just below their desired location and then a *shrink wrap* modifier was applied to the grid and targeted at the occlusal surface of the dentine. This caused the text object to deform and follow the contours. Finally, the position of the deformed text shape was fine-tuned to ensure that it overlapped the ADJ evenly.

To create the access, a cylinder of the same size as the original circle curve was placed within the scene and a *Boolean* difference applied to both the dentine and enamel to cut out the access.

Activity 1.2

To create Activity 1.2 (See Figure 7.2), the low-polygon tooth model from activity 1 was duplicated. The same process was followed with regard to extruding text and shrink-wrapping a bound grid to the dentine, however, because the exercise presents a a square-edged folded rectangular cut away across the two surfaces, the text was not configured to follow a circle curve. Instead, 6 straight-line extruded instances of the text were created, one for each of the short-sides of the rectangle and two each for the long sides. Splitting the long sides into two text objects meant that the text did not need to be extensively deformed to wrap around the corner at the disto-occlusal ridge.

Activity 1.3

To create Activity 1.3 (See Figure 7.3), activity 1.1 was again duplicated so that a region of unsupported enamel could be restored to the access created for activity 1.1. To create the region of unsupported enamel, the enamel and access cavity cylinder were duplicated in place. A *Boolean* intersection modifier was then applied to the duplicated shapes. This resulted in a 'puck' shape of enamel that exactly filled the hole left by the access cylinder. By resizing the duplicated cylinder, an unsupported section of enamel could be created of the desired size.

Activity 2.1

To create Activity 2.1 (See Figure 7.4), a cube was added to a scene containing the same tooth model as used in the previous exercises. This was sized and positioned to cover approximately a quarter of the crown of the tooth. A *Boolean* modifier set to difference was applied to the dentine and enamel objects to create the cut out. The cube was then marked as hidden. Next a cone was added and placed to represent caries spreading through the enamel. This was intersected with the enamel so that it conformed to its vertical extents and respected the contours of the internal and external surfaces. To create the ADJ caries, a 2D circle of the size of the desired lateral spread was positioned just below the enamel. A *shrink wrap* modifier was added to this and targeted at the enamel. The modifier was then applied to lock-in the deformation and finally, an extrude surface applied vertically in a negative direction to give the caries the desired depth.

Finally, a sphere was added to the scene and then cut in half via a loop-cut. The hole left by the cut was then filled in via the Grid Fill tool. This shape was scaled and positioned to represent the deeper caries. To not overlap the boundaries of the other shapes, the deeper caries received a *Boolean* difference modifier for each of the other tissues in the activity.

Where necessary, all shapes were differenced with the cube added initially to respect the cut-away design.

Activity 2.2

To create Activity 2.2 (See Figure 7.5), a 'reductive' approach was used. Real-world caries is the demineralisation of existing tissue and its resultant break down. In order to create an effective carious spread for this activity, this natural process was simulated by segmenting the existing tooth model and relabelling/colouring the subsections in-place to represent the decayed tissue - similar to if they had been demineralised by the natural process.

To create a stepped gradation of caries, a series of concentric spheres were added to the scene. The outermost sphere represented the overall spread of the caries and the mildly affected dentine. Successively smaller spheres denoted moderately and highly demineralised tissue respectively. These were *Boolean* intersected and differenced with a duplicate of the previous step to create a series of 'bowl' shaped regions. Where these bowls intersected with the tooth tissue, they were *Boolean* difference'd with it to constrain their size to respect the other object's dimensions.

Finally, a cone-shaped object was placed into the enamel to represent the initial ingress of the caries into the tooth. This repeated the process used above and linked up the outer surface with the initial infected tissue at the ADJ. A 'collar' of whiter demineralised enamel was also added to represent this feature in the carious lesions. This was achieved by adding a larger duplicate of the shape and subtracting the original to leave a hollow cone wrapping the original ingress point.

Activity 3.1

To create Activity 3.1 (See Figure 7.6), activity 2.2 was duplicated. The first step was to inverse the difference, such that rather than a quadrant being removed from the tooth to present a cut-out, instead just one quadrant was retained . To create the steps of guided caries removal detailed above, 3 concentric cylinders were added over the caries in the tooth. These were split horizontally so that 6 'stacked' objects were present in a 3x2 arrangement. The dividing line between the two rows was positioned such that the top row was located in the enamel and the bottom row within the dentine. All objects were intersected with the cube representing the visible quadrant, the top row of objects were intersected with the enamel, and the bottom row intersected with the dentine. The two outermost (relative to the centre of the tooth) cylinders for each row were also Boolean Modifier difference'd with the preceding steps so that the same tissue was not instructed for removal twice. These regions sub-divide the tooth such that addressable objects are present for the attachment of the pop up messages at their respective removal thresholds as described above.

Activity 3.2

To create the final activity, 4 separate Blender files were created, each with a separate quadrant of the tooth. These 4 separate files will be re-integrated as a single exercises in the simulated activity, however, designing them in separate files was more performant. Each quadrant was digitally prepared using Boolean modifiers to remove tissue in such a way as to produce the following states:

1. Quadrant 1 represents a preparation where caries at the ADJ has been correctly addressed (See Figure 7.7). A cylinder with a bevelled edge was created which represented the ideal material removal then a Boolean difference modifier applied to the other tissues to result in an ideal preparation.

2. Quadrant 2 presents a preparation exhibiting an improperly cleared ADJ (See Figure 7.8). The same preparation as used in the previous exercise was scaled to retain caries at the ADJ and Boolean difference modifiers applied to create the preparation. The tissue that should be removed to correct the shortcomings was separated via a separate boolean intersection so that informational messages can be attached at threshold percentages.

3. Quadrant 3 presents a preparation where unsupported enamel has been retained (See Figure 7.9). This exercise was created by removing the infected dentine and retaining the enamel above. The remaining enamel from this had an undesirable appearance, so a cone shape was intersected with the enamel to provide a better representation of what might be encountered if this error had been made. Again, the areas that should be removed to correct the error were segmented so that notifications could be applied later.

4. Finally, Quadrant 4 presented a preparation where deeper caries had been addressed prior to the addressing the caries at the ADJ (See Figure 7.10). This exercise used a scaled version of the ideal preparation from quadrant 1 and applied Boolean Modifiers to remove a column of tissue that fully cleared the pulpal floor and retained caries at the ADJ. Once more, the necessary required were segmented for later use.

E.2 Creating the Simulator Exercises

Once the design for each model of each exercise was complete, models were checked for bad geometry and ensured to be manifold and then each tissue was exported separately as a Stereolithography file (.stl). Each exercise was then created in turn using Virteasy Solid Editor (Laval, France) where each model was imported in turn and and the corresponding tissue type and colour assigned. Exercises were deployed to the simulators and acceptance testing conducted to ensure adequate performance was delivered.

Appendix F

Comparative Study Protocol

This appendix contains the protocol, developed following the SPIRIT guidelines, as used for the comparative study presented in Chapter 8.

A Randomised Comparative Study of 1st Year Undergraduate Dental Students for the Transfer of Cognitive Skills Acquired from Part-Task and Whole-Task Virtual Reality Training Exercises

Trial Registration

N/A - Non-clinical Educational Intervention

Protocol Version

Issue Date: 05 December 2023
Protocol Amendment Number: N/A
Author(s): AT, NM, SM, CWS

Date	Changes
2023-12	Original

Funding

Funding for this research has been provided by HRV (Laval) and The School of Clinical Dentistry, University of Sheffield.

Investigator

Mr Ashley Towers, Staff PhD Candidate, School of Clinical Dentistry, The University of Sheffield

Supervisory Team

Professor Nicolas Martin,
Professor of Restorative Dentistry, School of Clinical Dentistry, The University of Sheffield

Dr Steve Maddock,
Senior Lecturer, Department of Computer Science, The University of Sheffield

Professor Christopher Stokes,
Professor of Dental Education, School of Clinical Dentistry, The University of Sheffield

Role of Sponsor

HRV have provided part-funding to support this research and technical services to add software functionality to meet the research goals. HRV are not involved in the design or goals of this specific study but will be informed of the outcomes which may influence the development of future simulator functionality.

Introduction

Background and rationale

Virtual Reality (VR) simulators for pre-clinical dental education are currently designed to provide learners with whole-task simulations of dental operative procedures. Learners are able to practise performing these operative tasks in preparation for transferring the skills acquired to other simulation contexts or clinical practice.

However, the theory of motor skill specificity suggests that the reliability of which skills developed in a VR environment can be transferred to other contexts is limited because motor skills tend to be bound to the context they were acquired in and closely associated with the input signals received during training (Proteau, 1992). Any changes to these sensory inputs diminishes the reliability of transfer and means that improvements in performance as a result of the practice may only improve performance at the training task itself (Schmidt and Lee, 2014, pp. 218-219). To justify the time and effort expended in a training environment, it is vital that skills developed are able to transfer to other contexts so approaches which enhance transfer should be explored.

Dental operative skills are a complex combination of motor performance with a significant underlying cognitive component. The greater the complexity of a task and the more sub-goals it consists of, the more likely it is that it will exceed the processing capacity of a novice to perform it (Paas and Van Merrienboer, 1994). Part-task activities are one way in which this complexity can be managed by focussing the learner's attention on the aspects that facilitate constructing an understanding of the task and minimising energies wasted on task-irrelevant details. Using a part-task approach should reduce the time and mental effort required to acquire a skill than would be required when using conventional approaches (Paas and Van Merrienboer, 1994).

To identify candidate part-task exercises to explore this concept in a dental context, earlier work (TUoS Ethics Ref: 045165) performed a Task Analysis of the operative task of caries removal and cavity preparation. The output of this analysis was a structured Task Description which serves as a 'blueprint' for instruction and facilitates improvements to teaching through the explicit identification of all information that must be conveyed (Schmidt & Lee, 2014, p. 162). From this, a series of part-task exercises were developed for use with the School's VR simulators to provide focussed instruction on the "Establish Caries Free Margin at ADJ" goal of the operative procedure of caries removal.

These new exercises are the focus of this study and are intended to develop the underlying cognitive skills and declarative knowledge base that underpin the task. This is achieved by breaking the task down into smaller simpler tasks which focus on specific task elements and providing context-aware instructional messages to guide attentional focus towards key attributes. This cognitive rather than motor-skills focus to simulation should lead to improved transfer whilst avoiding the limitations caused by motor skills specificity. Additionally, the context-aware messages should focus the learners attention on the specific aspects of performance which should be attended to at each stage. Knowing where/when to focus a performer's limited reserve of attention is known to be a clear discriminator in expert performance (Schmidt and Lee, 2014, p. 64). If these exercises achieve their aim they should promote greater skill acquisition which can be demonstrated in other contexts.

This study seeks to expose two groups of participants to different educational interventions (bespoke part-task exercises and conventional whole-task training) designed to teach the steps and outcome criteria of the operative task of establishing a caries-free margin at the ADJ within an occlusal carious lesion.

Aim

Determine if a series of cognitive-focussed part-task VR exercises result in enhanced transfer of operative knowledge when compared to traditional whole-task VR training approaches.

Objectives

- Expose two equal groups of 1st year Dental Undergraduates to a VR educational intervention based on either part-task (experimental group) or whole-task (treatment-as-usual) approaches
- Measure performance of participants on a series of transfer tests which:
  - Measure how many of the attributes of performance were retained from the material covered during the intervention session via a Retention Activity.
  - Measure the ability to apply the retained knowledge via a Ranking Activity
  - Measure the retention of procedural knowledge via a Procedural Activity
- Perform statistical analysis of the data to identify if a significant difference is present between the groups that can be attributed to the intervention.
- Capture participant opinion and explore themes around the face-validity and perceptions of the intervention via a semi-structured interview.

Trial Design

This study is designed as a transfer test using a parallel group, two-arm, superiority trial with a 1:1 allocation ratio. The arms of the study are defined as an experimental group and a Treatment-as-Usual group.

Methods: Participants, interventions, and outcomes

Study setting

This study will be carried out at the School of Clinical Dentistry, University of Sheffield. Educational intervention sessions will be carried out in the on-site VR Simulation suite (Room C30). Follow up transfer test sessions will be carried out in a suitable on-site meeting room.

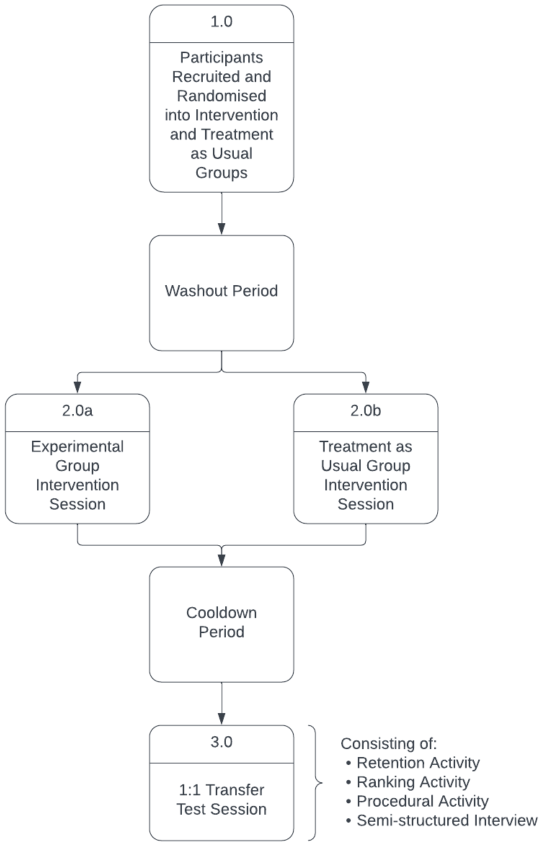
Eligibility criteria

Participants volunteering for the study must comply with the following at randomisation:

- Be enrolled in the 1st year of an undergraduate dental programme at the University of Sheffield
- Have completed the VR Suite familiarisation course.
- Must not hold an existing dental qualification (for example a qualified Dental Hygienist/Therapist or Dental Nurse seeking to extend their scope of practise).
- Must be on their first attempt (and not resitting) the 1st year of their respective programme.
- Be available to attend the intervention and transfer test session on the agreed dates.

The timing of the study ensures that participants are not undertaking any content in their degree programme that would confound the goals of this study.

Interventions



Eligible participants will be randomised in equal numbers to the experimental or treatment-as-usual group. Each group will undertake one intervention workshop session consisting of content appropriate to their arm of the study followed by, after a cooling-off period, a 1:1 transfer test session.

The intervention session for both arms of the study will take place in the simulation suite at the School of Clinical Dentistry which contains 10x Virteasy (HRV, Laval) VR Dental Skills Trainers. The learning objectives of the intervention sessions will be aligned to the "Clear the ADJ" goal and tasks from the preceding task analysis study (TUoS Ethics Ref: 045165).

The power calculation (see Sample Size below) shows that the number of participants required for this study exceeds the capacity of the simulation suite so multiple sessions will be provided. All participants present in each session will be from the same arm of the study. Participants will be asked to respect the confidentiality of their fellow participants and help avoid ‘pre-waring’ other participants by only discussing experiences relating to their participation within the group present until the conclusion of data gathering.

On rare occasions, participants may encounter discomfort during the use of the VR simulator. If this occurs or a participant is otherwise unable to complete the intervention session, they will be given an option to reschedule any uncompleted tasks separately within 48 hours. Failure to do so will require that they withdraw from the study.

The 1:1 transfer session for both arms of the study will take place in an appropriate meeting room at the School of Clinical Dentistry. In this session participants will attempt to apply knowledge acquired during the intervention to a series of tests. The session will conclude with a semi-structured discussion where participants explore themes around the acceptance and face-validity of the intervention.

Intervention Session: Treatment-as-usual Group Participants assigned to the treatment-as-usual group will undertake a 1 hour training session in the simulation suite. This will commence with a tutor-led presentation (with supporting Powerpoint slides) to introduce the procedure for caries removal at the ADJ followed by the use of existing simulation exercises to apply the content of the presentation. The learning objectives and material covered in the Treatment-as-usual session will be equivalent to that of the experimental group but will use whole-task simulation exercises instead of the focussed part-task exercises used by the experimental group.

Participants will be free to ask questions and seek clarifications. A note will be made of the participant and the question asked.

Intervention Session: Experimental Group Participants assigned to the experimental group will undertake a 1 hour training session in the simulation suite. The session will commence with a short introduction to the task followed by interaction with the part-task exercises developed for this study. Each of the exercises in this series presents an activity aligned to acquiring knowledge which contributes to the learning outcomes.

Participants will be free to ask questions and seek clarifications. A note will be made of the participant and the question asked.

Transfer Test Session After a cool down period, participants will attend a transfer test session where they will undertake a series of exercises to measure various aspects of the intervention. Both groups will undertake the same transfer test activities. All tests are designed to test knowledge related to the tasks of the “Establish Caries Free Margin at ADJ” goal from the task analysis. At the beginning of the session participants will be reminded that they are free to withdraw from the study at any point and will be asked to confirm if they have undertaken any further study relating to caries removal (privately or as part of their course)

Retention Activity The aim of this activity is to measure the participants retention of declarative facts. Participants will be presented with (n=5) 3D printed tooth models. Each of these models will contain a carious lesion which has been prepared in anticipation of a restoration being placed. Participants will be asked to take each tooth in turn and identify all issues with the state of the preparation based on the content covered during the intervention session. Participants will be provided with access to a dental probe to assist them in their evaluations.

Ranking Activity The aim of this activity is to assess the participants judgement and application of the retained knowledge. Participants will be asked to rank the five teeth that they have just critiqued into order of clinical acceptability. The most appropriately prepared tooth is the one they consider to be most suitable for immediately progressing to the next stage of treatment, the least appropriate is the one they consider to have the most egregious issues. Once they have completed the activity, they will be asked to explain the reasoning they used.

Procedural Activity The aim of this activity is to assess the participants retention of the procedural steps (in order) required to establish a caries-free margin at the ADJ. Participants will be provided with a tooth model (different from the five from previous activities) with a partially-prepared cavity and will be asked to describe which steps they would take (and the instruments they would use) in order to complete the preparation in the model.

Semi-structured Interview The aim of the semi-structured interview is to capture participant opinions on the intervention that they received. Face validity is an important concept in educational approaches and can impact upon engagement with the material. Responses to these questions will be thematically analysed in order to identify common perceptions of the approach

Outcomes

The purpose of this study is to explore if a series of VR part-task training exercises are superior to traditional VR training approaches when measured with a test of transfer to an alternative context. Transfer is an important test for educational interventions because the context in which the skills are acquired and practised often differs from the context in which they will be applied and it is known that certain approaches to practise are more or less effective than others (Schmidt and Lee, 2014, p.228). Therefore the following measures were selected to assess transfer so that any differences between the two arms of the study can be identified.

Primary outcome measures

- Retention Activity Score: Count of the number of attributes correctly/incorrectly identified in 3D printed models of carious/prepared teeth.
  - Analysis to establish presence of a statistically significant difference between groups.
- Ranking Activity Score: Sum of the squared-difference between participant determined ranking and a predetermined/expert defined correct ordering (the rational key). Superiority is the lowest difference between a participant's ordering and correct ordering.
  - Analysis to establish presence of a statistically significant score difference between groups.
  - Measure of concordance between groups, low concordance implies greater degree of randomness in responses or less well informed choices.
- Procedural Activity Score: Score produced from a recording of participant description of the next steps required to complete the establishment of a caries-free margin at the ADJ of a partially prepared example 3D printed model. Assessors will be blinded to the intervention received when producing a score based on a standardised marking scheme/rubric derived from the relevant goals from the task description.
  - Analysis to establish presence of a statistically significant difference between groups in the scores received.

Secondary outcome measures

- Gather participant opinions from a semi-structured interview to explore any themes relating to the acceptability and face validity of the intervention.

Participant timeline

Due to the different start dates of the two undergraduate programmes and the planned timing of data collection, participants enrolled on the BDS programme will be approached in the first instance. If insufficient BDS-enrolled participants are recruited, participants on the later-starting DH&DT programme will be approached separately. The timeline will remain equivalent but with dates relative to the equivalent sessions in the DH&DT timetable.

- In the week prior to session 2 of the timetabled VR Suite familiarisation course (w/c 15th of January 2024) an email will be sent to invite participants to the study providing: An introductory explanation, a participant information sheet and contact details for questions. The email will contain a link to a Google form to volunteer to participate.
- At the conclusion of each session 2 of the VR Suite familiarisation course (from w/c 22nd January 2024), students will have their attention drawn to the previously sent email and the study's aims verbally reiterated. Potential volunteers will be offered a chance to ask any questions prior to volunteering. NB participants will not be asked to sign up immediately, this is simply to highlight the research taking place within the School given the likelihood it could have been missed due to the high volume of emails students receive.
- Upon receipt of a completed volunteer form, participants will be assigned to a group following the randomisation schedule below and contacted via email to arrange their intervention session appointment time/date. Three session dates will be available for each arm of the study to present a greater chance of participant availability.
- A washout period of 3 weeks will be required between the taught SIMS session and the intervention therefore the first participants will become eligible to take part in the intervention session of the study during w/c 19th February 2024.
- On the agreed date, participants will attend a 1-hour intervention session in the simulation suite consisting of:

  Experimental Group:

  - Participants arrive and will be welcomed and directed to sit at a simulator. The simulators will have a post-it note attached displaying the Participant's ID.
  - Participants handed a copy of the Participant Information Sheet and asked to review the information and ask any questions
  - Once verbally confirmed they are happy to take part in the study, handed a Consent Form and asked to sign.
  - Session commences:
    - Introductory Powerpoint Presentation covering:

            · Request to respect the confidentiality of other participants
            · Introducing the learning objectives of the session and the exercises that will be used.
            · Participants will be informed that they can ask any questions or have material repeated at any point whilst interacting with the exercises
        * Participants interact with each of the newly developed part-task exercises following the instructions displayed on-screen. Each exercise focuses on a single aspect of the procedure. Approximately 10 minutes will be provided per exercise. Participants will be informed when to move to the next exercise.
        * Any questions asked by the participant will be recorded on the Participant Question Log
    – Session Conclusion:
        * Participants will be asked to consult their availability to arrange the transfer test session for approximately 7 days later. Any participants able to confirm their availability at the time will be emailed a calendar invite. Any who are not able will be contacted within 24 hours to arrange their appointment.
        * Participants will be thanked for their participation and their contribution to the research project so far.

Treatment as Usual Group:

- Participants arrive and are welcomed, then directed to sit at a simulator. The simulators will have a post-it note attached displaying the Participant's ID.
- Participants handed a copy of the Participant Information Sheet and asked to review the information and ask any questions
- Once verbally confirmed they are happy to take part in the study, handed a Consent Form and asked to sign.
- Session commences:
    * Powerpoint Presentation covering:
        · Request to respect the confidentiality of other participants
        · Introducing the learning objectives of the session
        · An explanation of the process and instruments used for caries removal at the ADJ including text and images.
        · Participants will be informed that they can ask any questions or have material repeated at any point during the session
    * Participants guided through an attempt at establishing a caries free margin using a whole-task simulation of a carious tooth

    - Participants to attempt a 2nd exercise independently (support provided if requested)
    - Any questions asked by the participant will be recorded on the Participant Question Log
  - Session Conclusion:
    - Participants will be asked to consult their availability to arrange the transfer test session for approximately 7 days later. Any participants able to confirm their availability at the time will be emailed a calendar invite. Any who are not able will be contacted within 24 hours to arrange their appointment.
    - Participants will be thanked for their participation and their contribution to the research project so far.

- On the date of the transfer session, participants will attend a 1 hour 1:1 session consisting of:
  - Participant arrives and is welcomed and thanked for returning for the 2nd part of the study. Participants are reminded that their participation is voluntary and they are free to withdraw at any point without any negative consequences.
  - Attention is drawn to the recording equipment and that it is aimed at their hands only and will capture their interactions with the models and our discussions.
  - The participant is informed that it is the exercises and training that they received that is being tested, not them individually, they should simply do their best based on what they can remember from their training session.
  - The participant is asked to confirm if they have undertaken any further study (taught or private) since their intervention session relating to caries removal. If yes, ask what they have done.
  - Transfer Tests:
    - Participant completes the Retention Activity
    - Participant completes the Ranking Activity
    - Participant completes the Procedural Activity
  - Semi-structured interview
    - The participant is informed that the final part of the study is an opportunity to share their views on the instruction they received.
    - Semi-structured interviews will be conducted to explore areas relating to the acceptability of the intervention from a student perspective. Sample questions and themes explored are listed below (See Transfer Test Session: Interview Question Prompts).
  - Session Conclusion:
    - The participant is thanked for their time and arrangements made to issue the Amazon Gift Card.

Sample Size

As stated above (Participant Timeline) both BDS and DH&DT students are suitable to participate in this study. However, due to the timing of data collection and their respective Simulation Introduction sessions, BDS students will be approached in the first instance:

With a sample population (BDS only) of 70 students and a 95% confidence interval, sample sizes of 18, 27 and 41 are required for confidence intervals of 20, 15 and 10 respectively. 41 participants is in excess of 50% of the cohort and previous experience has suggested that this level of recruitment is unlikely to be achievable. Therefore a target of 27 at the 15 confidence interval is set.

If insufficient BDS students are required to meet the above sample size, the cohort of 24 DH&DT students will be invited to participate. Including this cohort increases the population to 94 1st year students. Consequently the sample sizes increase to 19, 30 and 48 for the same confidence intervals. If DH&DT students are approached, the target sample size will increase to 30.

The operative task chosen for study is within the scope of practice of both programmes and the timing of the data collection would place both sets of students at the same relative point in their course, guarding against any confounding effects of different levels of relevant knowledge at the data collection stage.

Recruitment will continue to sample sizes in excess of the targets stated above to allow maintaining the confidence level of 15 in case of participants drop-out or to provide an improved confidence level.

All sample size calculations were performed with the online tool available at: `https://www.surveysystem.com/sscalc.htm` using the formula listed at: `https://www.surveysystem.com/sample-size-formula.htm` including corrections for finite populations.

Recruitment

Participants will be invited to the study via email during the w/c 15th January 2024. Due to the volume of emails that students receive, their attention will also be drawn to the message verbally and a brief outline of the research provided at the end of their timetabled SIMS Introduction session. Potential participants will be offered an opportunity to informally ask any questions.

If the target sample size has not been reached by Friday 9th February, a follow up email will be sent reminding participants of the study and offering a final opportunity to take part if desired.

If recruitment is still below the target above, recruited participants will take part in the study as described above. However, the separate round of recruitment from the DH&DT programme will be triggered and undertaken when this cohort reaches the equivalent point in the curriculum in June to early July 2024 (timetable TBC). The same participant timeline apply to this group.

Methods: Assignment of interventions

Sequence generation

Upon receipt of an intention to participate in the study, participants will be randomly assigned to one of two groups with a 1:1 allocation as per a computer generated randomisation schedule.

Allocation concealment mechanism

Participants will be allocated to groups known as Group A and Group B recorded on a private Google Sheet held on a restricted-access Shared drive. It will not be known if Group A or Group B will be the intervention or treatment-as-usual group at the time of assignment.

Implementation

The allocation sequence will generated using the online tool available at: `https://www.graphpad.com/quickcalcs/randomize1/`

The PI will enrol participants and assign them to the two groups as per the allocation sequence in the Google Sheet. The allocation of each group to an arm of the study will be performed at the conclusion of recruitment via a coin toss.

Blinding (masking)

The nature of the study and students' prior exposure to the simulation environment means that they will be aware (or can strongly infer) if they are receiving the treatment-as-usual intervention.

Prior to any analysis or evaluation of participant responses, responses will be collated and shuffled. Outcome assessors and data analysts will be blinded to the intervention received by each participant.

Methods: Data collection, management, and analysis

Data collection methods

The 1:1 Transfer Task session will be recorded (audio and visual stored as H.264 MOV) on two redundant secured and encrypted recording devices. This will capture the audio of the participant's answers and any discussion. The camera will be pointing at the participants' hands to capture their interactions with the Transfer Task Models. A PostIt note containing the Participant ID will be affixed to the table in frame of the recording. This recording will be transferred on to a University storage device approved for research as soon as possible after the session. The recording will be used, as described below, to transcribe the participants' responses and answers for further analysis. These recordings will be destroyed at the conclusion of the research project.

Retention Activity Participants will be provided with (n=5) 3d printed models of teeth, each of which exhibit a combination of the positive and negative attributes covered in the intervention session to varying degrees. These models will range from a clinically acceptable (i.e. ready for immediate restoration) to those with iatrogenic damage. Each model will be labelled with a geometric shape identifier (to avoid anchoring or any implied order) and placed in a random group on the table. Participants will be instructed to take each tooth in turn and verbally assess each model, drawing on the content of their intervention session to note the attributes present. Participants will have access to a dental probe to optionally assist in their exploration.

Later, the recordings will be reviewed and the attributes as identified by the participant will be transcribed. Where there is ambiguity in the participants response, a clinical member of the research team will adjudicate if the description as-recorded is adequate to have identified the described feature. To mitigate for bias, the assessor will be blinded to the intervention received during transcription.

If any errors are made in the identification of attributes (false positive) these will be recorded separately and subjected to separate analysis.

Assessment and the review of progress is a key part of the overarching procedure of caries removal and cavity preparation. Furthermore, reflective practice is a requirement of all dental professionals. Being able to recognise and identify deficiencies in a preparation is a natural first step towards corrective action so any lack of awareness of these attributes is considered a negative indicator to the acquisition of the requisite skills.

Ranking Activity After assessing each tooth model individually, participants will be asked to take the 5 teeth and use their judgement to rank them from best to worst (in terms of the attributes covered in the intervention session). The participant's determined ranking will be captured by the recording device and transcribed at a later date for further analysis.

Judgement and the weighing up of competing factors is a core skill necessary for determining subsequent actions. A good understanding of the desirable and undesirable attributes demonstrated in the tooth models will be revealed by correctly ranking the attempts from best to worst in agreement with the rankings of a clinician. This test compares with similar judgement tests carried out elsewhere in the dental undergraduate programme.

The use of 5 models in the above tests was selected because this number is the lower limit of the quantity of items that can be stored in working memory (Miller, 1956). The participants' working memory will already be taxed whilst attempting to apply relatively new concepts from the intervention session. Therefore, it was considered that compounding the reduction of the available germane cognitive load may confound the goals of study by shifting the task towards one of managing cognitive load rather than that of assessing and evaluating attributes.

Procedural Activity Participants will next be provided with a part-completed preparation in a 3d printed tooth model (different from the previous 5 models) and will be asked to describe the steps required to complete the clearing of the ADJ including the instrumentation they would choose. Their recording will be reviewed at a later date and marked against a standardised marking sheet which will assess the extent of the deviation from the correct procedure taught in the sessions derived from the task description.

This form of OSCE style assessment has high face-validity and is an approach used widely within the dental undergraduate degree.

Semi-structured interviews The face-validity of the approach to undergraduate students is of interest for evaluating this educational approach. Therefore, participants who are assigned to the experimental arm of the study will be invited to share their opinions of the intervention in comparison to the style of exercises they encountered during their timetabled simulation introduction sessions via a semi-structured interview. Participants who are assigned to the treatment-as-usual approach will be asked questions relating to how they feel the simulation experience could be improved to assist with skills acquisition. Prompt questions will be asked but the format of the interview will not be too prescriptive to promote free discussion along themes that are of interest and relevance to the participant.

The recordings of these interviews will be reviewed at a later date and subjected to thematic analysis using the approach presented in Burnard et al (2008).

Plans to promote participant retention

To ensure the inclusion of participants, they will be offered an incentive of a £20 paid in gift vouchers for completion of the two sessions. To improve the likelihood of attendance at the 1:1 transfer session this will be scheduled in-person at the end of the intervention session. If a participant is unable to commit to a time at the end of the session they will be contacted by email the next working day with some suggested dates.

Plans for data security and storage

All data will be stored on a Google Drive created specifically for this research project using a University managed Google account. Only members of the research team will be granted access to this drive using their University managed accounts.

All data provided by participants will be stored in files and folders named using the participant ID. Whilst the filenames and folders will use the participant ID, members of the research team may be able to identify the participant by their voice or likeness in the video recordings. No audio or visual manipulation of the capture will be performed to anonymise the participants in this recording whilst stored for the duration of this project. The video recordings will not be shared outside of the project team.

A spreadsheet stored in a separate Google Drive accessible only to the researcher who is conducting the one-to-one sessions will contain the link between the personally identifiable information and the participant ID.

All video recordings and the document linking participants and their anonymous identifiers will be destroyed at the end of the researcher's PhD project (currently, November 2025 at the latest).

Statistical methods

The following statistical tests will be applied to the above results and their implications discussed in any publications arising from this work. Tests have been selected to explore transfer task performance within and between the arms of the study. In all cases the null hypothesis is that there is no difference between the two groups:

Primary Outcomes

Retention Activity Statistical tests for the Retention Activity will consider if there is a statistically significant difference between the means of the two study groups. Tests will measure the retention of declarative facts in relation to their application to a transfer task to establish if retention is superior using the experimental approach. Superiority is defined as correctly identifying a greater number of attributes.

Overall Retention The average total number of attributes identified across all models by each participant will be compared and a poisson-rate ratio test performed to explore if there is a statistical relationship between the total number of features identified and the intervention received.

Total Correct The average total number of attributes identified correctly will be compared and a poisson-rate ratio test performed to explore if there is a statistical relationship between the total number of features identified and the intervention received.

Total Incorrect The average total number of attributes identified incorrectly will be compared and a poisson-rate ratio test performed to explore if there is a statistical relationship between the total number of features mis-identified and the intervention received.

Comparison of attributes across models To explore if either intervention favours knowledge acquisition of particular attributes of performance, identification of each attribute will be calculated as a percentage of the maximum identifiable across all models. For example if a given attribute is present on 4 of the models and a participant correctly identifies it on 3 of the models, a score of 75% will be recorded.

Independent samples T-Tests will be performed for each of the attributes to establish if there is a difference of means between the two groups for any of the attributes. NB tests are intended to be analysed and discussed independently - but a Bonferroni correction will be applied before rejecting the null hypothesis if any of the tests are combined.

Ranking Activity For each arm of the study, the mean ranking and a pairwise ranking comparison of each model will be calculated using the method described in Finch (2022). A lower value for the mean ranking indicates that the item is preferred by members of the study group (where 1 is the highest rank). Likewise, a pairwise comparison will produce a matrix indicating how frequently any given item was preferred over another. After calculating these initial descriptive statistics, Kendall's Coefficient of Concordance (also known as Kendall's W) will be calculated to explore the degree of agreement of ranking within each group.

The above statistics will permit discussion of the impact of the intervention received and how consistent its impact was across participants within the groups. A low agreement within groups could suggest that rankings were less informed or had a random nature. Narrative discussion may also result from specific model rankings for example where a particular model is ranked very differently between the two arms of the study.

The above analysis measures the agreement within each arm of the study without reference to any correct ordering. Therefore, each participant's ranking will be compared to a rational scoring key (nominally the *correct* order) as determined by subject matter experts to produce a score. The score will be derived from the sum of the distance-squared of each ranked item when compared to the reference ranking. Squaring the difference more heavily weights the rankings of items with the greatest deviation from the rational scoring key (Legree et al. 2005).

A Two Sample t-test for difference in means will then be applied to the scores of the two groups to explore if there is a statistical difference between the scores which can be attributed to the intervention received.

Procedural Activity The scores received by participants against the standardised marking sheet will be compared via (assuming normally distributed data) an independent samples t-test between the intervention received to establish if there is a statistically significant difference between the two arms of the study in relation to their ability to explain the correct procedural steps to complete the clearing of the ADJ.

NB all tests above are intended to be analysed and discussed independently - but a Bonferroni correction will be applied before rejecting the null hypothesis if analysis leads to any of the tests being combined.

Secondary Outcomes

Semi-structured interview The semi-structured interview will explore emergent themes from participant impressions and will not be subject to statistical analysis

Methods: Monitoring

Data monitoring

As a small low-risk educational intervention a data monitoring committee is not required for this study. No interim results will be shared and the trial will only be terminated if participants are unable to complete any of the transfer tasks.

Harms

The equipment used for this study is in regular use as part of timetabled sessions and participants will, at the time of the study, have received instruction on its use. However, if any participant encounters any discomfort from the use of the simulator they will be offered the opportunity to withdraw from the study without consequence.

Auditing

All interventions will be conducted by the PI and data will be captured in a consistent manner. Results will be monitored by the supervisory team and subsequently defended as part of the PhD assessment process.

Ethics and dissemination

Research ethics approval

This protocol along with other supporting material will be submitted to the School of Clinical Dentistry ethical review process and an identifier assigned prior to any data collection. It is not anticipated that any changes will be required to this protocol during the duration of the study. However, should this expectation prove to be false, following discussions with the supervisory team, the Departmental Ethics Lead will be consulted to discuss amendments to the approved protocol.

If any substantial change requires amended consent from the participants, an updated consent form and participant information sheet will be subjected to the Ethical Review process. The updated form will then be shared with the participants who will be asked to re-consent under the modified conditions or to withdraw from the study and to have their data destroyed.

Consent or assent

A copy of the participant information sheet (PIS) and the consent form will be included with the recruitment email so that participants can review the information and ask any questions prior to volunteering. Additionally, at the beginning of the intervention session, participants will be provided with a physical copy of the PIS and given time to review the information and ask any questions. After this, participants will be asked by the PI to sign two copies of the consent form, one for their own records and one to be stored in the project file.

Confidentiality

All participant information will be stored on a Google Drive used for research data associated with this study. Access to this drive is restricted to members of the project team. Hard-copies of signed consent forms will be stored in a locked filing cabinet in the office of the PI. All data will be keyed exclusively to the participant identifier and the relationship between the individual and the identifier will be stored in a single location accessible to only the project team.

Declaration of interests

The PI is in receipt of part-funding for his PhD programme from HRV (Laval) however, the sponsor is not directly involved in the design or running of this study. Only the completed results will be shared at its conclusion.

No harm is anticipated from participation in this study so no provision has been made for post-trial care or compensation.

Dissemination policy

Results from this study will be published as part of the Investigators PhD thesis. The sponsor of the project is entitled to temporarily delay publication of the thesis if it is deemed to contain commercially sensitive information as per the research sponsorship agreement terms agreed with The University of Sheffield.

Authorship of any publications arising from this study will be shared between the investigator and the supervisory team. No professional writers will be used for any resultant publication beyond the normal editorial oversight of the publisher.

Anonymised participant data and a copy of this protocol will be shared in the appendices of the thesis and also in data-sharing statements as part of open access publications. Selected results may be submitted for publication in an appropriate journal with access to underlying anonymised data provided upon request if/when needed by the journal's publication requirements.

Appendices

Informed consent materials

All consent information will be provided via the Research Ethics Approval System in the exact format that it will be provided to participants. This information is not repeated in this document.

Biological specimens

Not applicable.

Appendix G

Analysis Source Code

G.1 Retention Activity

```r
rm(list=ls())
install.packages("irnNA", "readxl", "crank", "pmr", "ggpubr", "nortest",
    "epiDisplay")

library(readxl)
library(irrNA)
library(crank)
library(pmr)
library(PlackettLuce)
library(ggpubr)
library(nortest)
library(dplyr)


# Read in source file
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
source_file <- 'RetentionTestResults.xlsx'
retention_test <- read_excel(source_file, "Full Data")

# ---------------------------------------------------------------
# Establish if total correct data is from a Poisson Distribution using a
    MinChisq Goodness of Fit
# ---------------------------------------------------------------
# Null hypothesis: The numbers of correctly identified attributes are
    from a Poisson distribution
# Alternative hypothesis: The numbers of correctly identified attributes
    are different to a poisson distribution

empirical_data <- retention_test$`Total Correct`
empirical_data <- empirical_data[!is.na(empirical_data)] # Remove NAs
    caused by sub-totals at bottom of imported sheet
```

```r

gf <- goodfit(empirical_data, type= "poisson",method= "MinChisq")
summary(gf)


# ---------------------------------------------------------------------
# Test if total correctly identified attributes differs between the groups
# ---------------------------------------------------------------------
# Null hypothesis: The numbers of correctly identified attributes are equal
# Alternative hypothesis: The numbers of correctly identified attributes are different

group_a <- filter(retention_test, retention_test$Group == "A")
group_b <- filter(retention_test, retention_test$Group == "B")

cond_a <- group_a$`Total Correct`
cond_b <- group_b$`Total Correct`
mean(cond_a)
min(cond_a)
max(cond_a)

mean(cond_b)
min(cond_b)
max(cond_b)

poisson.test(x = c(sum(cond_a), sum(cond_b)), T = c(length(cond_a), length(cond_b)), alternative = c("two.sided"))

# ---------------------------------------------------------------------
# Establish if total incorrect data is from a Poisson Distribution
# ---------------------------------------------------------------------
# Null hypothesis: The numbers of correctly identified attributes are from a Poisson distribution
# Alternative hypothesis: The numbers of correctly identified attributes are different to a poisson distribution

incorrect_data <- retention_test$`Total Incorrect`

gf <- goodfit(incorrect_data, type= "poisson",method= "MinChisq")
summary(gf)

# ---------------------------------------------------------------------
# Test if total incorrectly identified attributes differs between the groups
# ---------------------------------------------------------------------
# Null hypothesis: The numbers of incorrectly identified attributes are equal
# Alternative hypothesis: The numbers of incorrectly identified attributes are different

inc_a <- group_a$`Total Incorrect`
```

```r
inc_b <- group_b$`Total Incorrect`

mean(inc_a)
min(inc_a)
max(inc_a)

mean(inc_b)
min(inc_b)
max(inc_b)

poisson.test(x = c(sum(inc_a), sum(inc_b)), T = c(length(inc_a), length(
    inc_b)), alternative = c("two.sided"))

# -------------------------------------------------------------------
# Test if total retained attributes differs between the groups (
    regardless of if correct/incorrectly applied)
# -------------------------------------------------------------------
# Null hypothesis: The numbers of retained attributes are equal
# Alternative hypothesis: The numbers of retained attributes are
    different

ret_a <- group_a$`Retained`
ret_b <- group_b$`Retained`

mean(ret_a)
min(ret_a)
max(ret_a)

mean(ret_b)
min(ret_b)
max(ret_b)

poisson.test(x = c(sum(ret_a), sum(ret_b)), T = c(length(ret_a), length(
    ret_b)), alternative = c("two.sided"))

# ------------
# Are there differences in correct minus incorrect?
# ------------

dif_a <- group_a$`Total Correct` - group_a$`Total Incorrect`
dif_b <- group_b$`Total Correct` - group_b$`Total Incorrect`

mean(dif_a)
min(dif_a)
max(dif_a)

mean(dif_b)
min(dif_b)
max(dif_b)
poisson.test(x = c(sum(dif_a), sum(dif_b)), T = c(length(dif_a), length(
    dif_b)), alternative = c("two.sided"))

# -------------------------------------------------------------------
```

```r
# Explore if there is any difference between the individual tasks e.g. does either arm
# perform better on specific tasks than the other (in terms of total correct/incorrect)
# ---------------------------------------------------------------------
# Null hypothesis: The performance on each individual task is equal for each group
# Alternative hypothesis: The performance on individual tasks is different between arms of the study

unimputed_test = read_excel(source_file, "Full Data")
group_a <- filter(unimputed_test, unimputed_test$Group == "A")
group_b <- filter(unimputed_test, unimputed_test$Group == "B")
partial_column_name = c("Diamond - Total", "Triangle - Total", "Square - Total", "Circle - Total", "Rectangle - Total")
correctness = c("Correct", "Incorrect")

for (status in correctness) {
  cat(" ########### ", status, " ########### \n")
  stddev_a = c()
  stddev_b = c()
  for(col in partial_column_name) {

    cat(">>>>>>> Analysis of : ", col, status, "\n")

    column <- paste(col, status)
    cond_a <- group_a[[column]]
    stddev_a = c(stddev_a, cond_a)

    cond_b <- group_b[[column]]
    stddev_b = c(stddev_b, cond_b)
    cat("Group A Mean:", mean(cond_a), "\n")
    cat("Group B Mean:", mean(cond_b), "\n")
    print(poisson.test(x = c(sum(cond_a), sum(cond_b)), T = c(length(cond_a), length(cond_b)), alternative = c("two.sided")))
    cat("------------------------------\n")
  }
  cat ("Standard deviation Group A ", status, ": ", sd(stddev_a), "\n")
  cat ("Standard deviation Group B ", status, ": ", sd(stddev_b), "\n")
  print(var.test(stddev_a, stddev_b))
}

# ---------------------------------------------------------------------
# Explore retention of attributes present across multiple tasks i.e. are participants likely
# to consistently identify a particular attribute and is there a difference between arms of the study?
# Binary proportion - see: https://statsandr.com/blog/fisher-s-exact-test-in-r-independence-test-for-a-small-sample/
# ---------------------------------------------------------------------
unimputed_test <- read_excel(source_file, "Unimputed Totals")
group_a <- filter(unimputed_test, unimputed_test$Group == "A")
group_b <- filter(unimputed_test, unimputed_test$Group == "B")
```

```r

columns <- c("ADJ Identified", "Unsupported", "Retained Floor", "Fissure
    Pattern", "Probe Use")
multiplier <- c(5,5,4,5,4) # Number of examples the attribute is present
    on. Used for calculating max
i <- 1
for (col in columns) {
  cat(">>>>>>> Analysis of : ", col, "\n")
  identified_group_a <- sum(group_a[[col]])
  identified_group_b <- sum(group_b[[col]])
  ommitted_a <- (multiplier[i] * length(group_a[[col]])) - identified_
    group_a
  ommitted_b <- (multiplier[i] * length(group_b[[col]])) - identified_
    group_b
  dat <- data.frame(
    "Identified" = c(identified_group_a, identified_group_b),
    "Ommitted" = c(ommitted_a, ommitted_b),
    row.names = c("Group A", "Group B"),
    stringsAsFactors = FALSE
  )
  colnames(dat) <- c("Identified", "Ommitted")
  cat("Data: \n")
  print(dat)
  cat("Expected frequencies (all should be >5)\n")
  print(chisq.test(dat)$expected)

  print(chisq.test(dat))
  i <- i + 1
}
```

G.2 Ranking Activity

```r
rm(list=ls())
install.packages("irnNA", "readxl", "crank", "pmr", "ggpubr", "nortest")
library(readxl)
library(irrNA)
library(crank)
library(pmr)
library(PlackettLuce)
library(ggpubr)
library(nortest)

# Read in source file
source_file = 'RankingActivityResults.xlsx'
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
group_a = read_excel(source_file, "A Transposed")
group_b = read_excel(source_file, "B Transposed")

# --- Calculate Kendall's W and other statistics

# Store results of each call to stats function in a list
l_irr_a = kendallNA(group_a)
```

```
l_irr_b = kendallNA(group_b)

# Create and label the data then calculate the differences between all statistics generated
irr = cbind(names(l_irr_a),data.frame(matrix(unlist(l_irr_a),nrow=length(l_irr_a), byrow=TRUE)),
                           data.frame(matrix(unlist(l_irr_b),nrow=length(l_irr_b), byrow=TRUE)))
names(irr) <- c("Measure","Group A", "Group B")
irr$Difference <- abs((irr$`Group B` - irr$`Group A`))
irr

# Compare when excluding outlier:
group_a_excl_outlier = group_a[ , -which(names(group_a) %in% c("Participant 26"))]
kendallNA(group_a_excl_outlier)
# --- Calculate mean ranks

# Deprecated to assign row names, but it works better with transpose than a labels column
row.names(group_a) = c("Diamond", "Triangle", "Square", "Circle", "Rectangle")
row.names(group_b) = c("Diamond", "Triangle", "Square", "Circle", "Rectangle")

mr_group_a = t(group_a)
mr_group_b = t(group_b)

meanranks(mr_group_a)
meanranks(mr_group_b)

all_mean_ranks = rbind(mr_group_a, mr_group_b)
all_mean_ranks.mean_ranks<-meanranks(all_mean_ranks)
all_mean_ranks.agg<-rankagg(all_mean_ranks)
all_mean_ranks.desc<-destat(all_mean_ranks.agg)

# --- Test for Random Mean
null_mean<-rep(3,5) # 3 is t+1/2 (t=5 items, 6 == number of items?)
A<-((12*27)/(5*(5+1)))
chi<-A*sum((all_mean_ranks.desc$mean.rank-null_mean)^2)
chi
dchisq(chi,4)

# --- Calculate Standard Deviations
dt_group_a<-data.frame(mr_group_a)
sd(dt_group_a$Diamond)
sd(dt_group_a$Triangle)
sd(dt_group_a$Square)
sd(dt_group_a$Circle)
sd(dt_group_a$Rectangle)

dt_group_b<-data.frame(mr_group_b)
sd(dt_group_b$Diamond)
```

```r
sd(dt_group_b$Triangle)
sd(dt_group_b$Square)
sd(dt_group_b$Circle)
sd(dt_group_b$Rectangle)

# -- Generate Pairwise table

# initialize blank results table
items = c("Diamond", "Triangle", "Square", "Circle", "Rectangle")
m = matrix(0, nrow=5, ncol=5)
pairwise_a = data.frame(row.names = items, m)
names(pairwise_a) <- items
pairwise_b = data.frame(row.names = items, m)
names(pairwise_b) <- items

# Loop over all permutations
for (row in items) {
  for (column in items) {
    if (column == row) {
      next # Skip where comparing with self
    }
    for(i in 1:nrow(mr_group_a)) {

      ranking <- mr_group_a[i,]
      # Get numeric value of iterator item(s)
      ivalr = which(row == items)
      ivalc = which(column == items)

      # Get the ranking position for the rater
      rankr = which(ivalr == ranking)
      rankc = which(ivalc == ranking)

      if (rankr > rankc) {
        pairwise_a[column, row] <- pairwise_a[column,row] + 1
      }
    }

    for(i in 1:nrow(mr_group_b)) {

      ranking <- mr_group_b[i,]
      # Get numeric value of iterator item(s)
      ivalr = which(row == items)
      ivalc = which(column == items)

      # Get the ranking position for the rater
      rankr = which(ivalr == ranking)
      rankc = which(ivalc == ranking)

      if (rankr > rankc) {
        pairwise_b[column, row] <- pairwise_b[column,row] + 1
      }
    }

```

```
  }
}
print("Group A - Pairwise")
pairwise_a
print("Group B - Pairwise")
pairwise_b

# --- Calculate Marginals Matrix comparisons
mr_group_a.agg<-rankagg(mr_group_a)
mr_group_a.desc<-destat(mr_group_a.agg)
print("Group A - Marginals Matrix")
t(mr_group_a.desc$mar) # Transpose to put items as rows and rankings as
    columns

mr_group_b.agg<-rankagg(mr_group_b)
mr_group_b.desc<-destat(mr_group_b.agg)
print("Group B - Marginals Matrix")
t(mr_group_b.desc$mar) # Transpose to put items as rows and rankings as
    columns

# --- Score vs Rational Scoring Key
dt_group_a$score<-(((dt_group_a$Diamond - 1)^2) + ((dt_group_a$Triangle -
     2)^2) + ((dt_group_a$Square - 3)^2) + ((dt_group_a$Circle - 4)^2) +
    ((dt_group_a$Rectangle - 5)^2))
dt_group_b$score<-(((dt_group_b$Diamond - 1)^2) + ((dt_group_b$Triangle -
     2)^2) + ((dt_group_b$Square - 3)^2) + ((dt_group_b$Circle - 4)^2) +
    ((dt_group_b$Rectangle - 5)^2))
mean(dt_group_a$score)
mean(dt_group_b$score)

# --- Test for normality
ttest_data.normality<-c(dt_group_a$score, dt_group_b$score)
hist(ttest_data.normality, col="steelblue", main="Normal") # Using
    Sturges' Rule for the number of bins
hist(ttest_data.normality, breaks = seq(min(ttest_data.normality), max(
    ttest_data.normality), length.out = 15), col="steelblue", main="Normal
    ") # Extend to help visualise
qqnorm(ttest_data.normality)
qqline(ttest_data.normality, col = "blue")
ggqqplot(ttest_data.normality)

shapiro.test(ttest_data.normality)
nortest::ad.test(ttest_data.normality)
nortest::cvm.test(ttest_data.normality)
nortest::lillie.test(ttest_data.normality)
nortest::pearson.test(ttest_data.normality)
nortest::sf.test(ttest_data.normality)

# --- T-Test
ttest_data<-data.frame(dt_group_a$score)
ttest_data$dt_group_b.score<-dt_group_b$score
names(ttest_data) <- c("Group A Score", "Group B Score")
t.test(ttest_data$`Group A Score`, ttest_data$`Group B Score`, var.equal
```

```
    = TRUE)

# --- As above but excluding outlier
group_a_excl_outlier = dt_group_a[row.names(dt_group_a) != "Participant 26", , drop = FALSE]
group_a_excl_outlier$score<-(((group_a_excl_outlier$Diamond - 1)^2) + ((group_a_excl_outlier$Triangle - 2)^2) + ((group_a_excl_outlier$Square - 3)^2) + ((group_a_excl_outlier$Circle - 4)^2) + ((group_a_excl_outlier$Rectangle - 5)^2))
mean(group_a_excl_outlier$score)
ttest_data.normality<-c(group_a_excl_outlier$score, dt_group_b$score)
shapiro.test(ttest_data.normality)
t.test(group_a_excl_outlier$score, dt_group_b$score, var.equal = TRUE)
```

G.3 Procedural Activity

```
rm(list=ls())
library(readxl)
library(dplyr)
library(ggpubr)
library(stringr)

# Read in source file
source_file <- 'ProceduralTaskResults.xlsx'
setwd(dirname(rstudioapi::getSourceEditorContext()$path))
procedural_test <- read_excel(source_file, "Sheet1")

# -------------------------------------------------------------------
# Establish if Overall Scores are normally distributed
# -------------------------------------------------------------------
# Null hypothesis: The scores are from a normal distribution
# Alternative hypothesis: The are not normally distributed

group_a <- data.frame(filter(procedural_test, procedural_test$Group == "A"))
group_b <- data.frame(filter(procedural_test, procedural_test$Group == "B"))
ttest_data.normality<-c(group_a$`Overall.Score`, group_b$score$`Overall.Score`)
# hist(ttest_data.normality, col="steelblue", main="Normal") # Using Sturges' Rule for the number of bins
# qqnorm(ttest_data.normality)
# qqline(ttest_data.normality, col = "blue")
ggqqplot(ttest_data.normality)

shapiro.test(ttest_data.normality)

# Test if there is a difference in means between the two groups
# --- T-Test Overall Score
ttest_data<-data.frame(group_a$`Overall.Score`)
ttest_data$group_b.score<-group_b$`Overall.Score`
names(ttest_data) <- c("Experimental", "TAU")
```

```
t.test(ttest_data$`Group A`, ttest_data$`Group B`, var.equal = TRUE)

boxplot(ttest_data, main = "Overall Score")


print("Overall Standard Deviations")
print("Overall Group A")
print(sd(group_a[[column]]))
print("Overall Group B")
print(sd(group_b[[column]]))
print(var.test(group_a[[column]], group_b[[column]]))

# Repeat the above for each of the domains

column_names = c("Selection.of.Bur", "High.vs.Low.speed", "Accessible.at.
    ADJ", "Infected.vs.affected", "Unsupported.and.access.to.ADJ", "
    Evaluation.and.Iteration", "Pulpal.Floor")

for (column in column_names) {
  cat(" ########### ", column, " ########### \n")
  title = str_replace_all(column, "\\.", " ")
  domain_ttest_data<-data.frame(group_a[[column]])
  domain_ttest_data$group_b.score<-(group_b[[column]])

  # Test for normality
  domain_ttest_data.normality<-c(group_a[[column]], group_b[[column]])
  print(shapiro.test(domain_ttest_data.normality))
  print(ggqqplot(domain_ttest_data.normality, title = title))

  print("Standard Deviations")
  print("Group A")
  print(sd(group_a[[column]]))
  print("Group B")
  print(sd(group_b[[column]]))
  print(var.test(group_a[[column]], group_b[[column]]))

  group_a_col_name = paste("Group A - ", column)
  group_b_col_name = paste("Group B - ", column)
  names(domain_ttest_data) <- c(group_a_col_name, group_b_col_name)

  print("===== Manually choose which is the most appropriate based on the
     Shapiro Test =====")
  # Perform both the t-test and Mann-Whitney test then choose the most
    appropriate by hand with reference to the normality above
  print(t.test(domain_ttest_data[[group_a_col_name]], domain_ttest_data[[
    group_b_col_name]], var.equal = TRUE))

  print(wilcox.test(domain_ttest_data[[group_a_col_name]], domain_ttest_
    data[[group_b_col_name]], exact=FALSE))

  # rename the columns to improve presentation for the box plots
  names(domain_ttest_data) <- c("Experimental", "TAU")

```

```r
  print(boxplot(domain_ttest_data, main = title))
  cat("------------------------------\n")
}
```

Appendix H

Comparative Study Data Extraction Form

The below form was used whilst reviewing the recordings of participants performance on the transfer task to capture attributes listed during the retention test and their suggested ordering for the ranking activity.

Participant ID ___

Model: ♦
Order: ____

Correct Attributes	Incorrect Attributes
ADJ ○ Successfully cleared ADJ. ○ Unsupported enamel addressed. **Pulpal Floor** ○ Pulpal Floor Caries Retained ○ Smooth/even base **Form** ○ Follows Fissure Pattern ○ Good angulation **Overall** ○ Intentionally used probe. ○ Good preparation for this stage	**ADJ** ○ Believes that caries is at ADJ. ○ Erroneously identifies unsupported enamel. **Pulpal Floor** ○ Suggests that pulpal floor caries should be removed. ○ Suggests floor is rough or uneven. **Form** ○ Suggests fissure pattern not followed. ○ Suggests that angulation is poor. **Overall** ○ Does not use probe at all/effectively. ○ Suggests preparation requires improvement.
Total: __/8	**Total: __/8**

Model: ▲
Order: ____

Correct Attributes	Incorrect Attributes
ADJ ○ Caries detected at ADJ. ○ Unsupported enamel detected. **Pulpal Floor** ○ Pulpal Floor Caries Retained ○ Smooth/even base **Form** ○ Follows Fissure Pattern ○ Good angulation **Overall** ○ Intentionally used probe. ○ Recognises stage is not complete.	**ADJ** ○ Believes caries at ADJ addressed. ○ Believes no unsupported enamel. **Pulpal Floor** ○ Suggests that pulpal floor caries should be removed. ○ Suggests floor is rough or uneven. **Form** ○ Suggests fissure pattern not followed. ○ Suggests that angulation is poor. **Overall** ○ Does not use probe at all/effectively. ○ Suggests stage is complete.
Total: __/8	**Total: __/8**

Model: ■

Correct Attributes	Incorrect Attributes
ADJ ○ Caries detected at ADJ. ○ Argues unsupported enamel present or absent (ambiguous). **Pulpal Floor** ○ Pulpal Floor Caries Retained. ○ Floor is rough and uneven. **Form** ○ Follows Fissure Pattern ○ Poor angulation identified. **Overall** ○ Intentionally used probe. ○ Recognises stage is not complete.	**ADJ** ○ Believes caries at ADJ addressed. **Pulpal Floor** ○ Suggests that pulpal floor caries should be removed. ○ Suggests floor is smooth/even. **Form** ○ Suggests fissure pattern not followed. ○ Believes angulation is acceptable. **Overall** ○ Does not use probe at all/effectively. ○ Suggests stage is complete.
Total: __/8	**Total: __/7**

Model: ●

Correct Attributes	Incorrect Attributes
ADJ ○ Caries detected at ADJ. ○ Unsupported enamel detected. **Pulpal Floor** ○ Pulpal Floor Caries addressed prematurely. ○ Excessively cleared pulpal floor (no affected retained) **Form** ○ Does not follow fissure pattern. ○ Good/acceptable angulation. ○ Margin requires refinement. **Overall** ○ Intentionally used probe. ○ Recognises stage is not complete.	**ADJ** ○ Believes caries at ADJ addressed. ○ Believes no unsupported enamel. **Pulpal Floor** ○ Suggests that pulpal floor caries was appropriately removed. ○ Suggests an appropriate amount of caries removed. **Form** ○ Suggests fissure pattern was followed. ○ Suggests that angulation is poor. ○ Suggests margin is well prepared. **Overall** ○ Does not use probe at all/effectively. ○ Suggests stage is complete.
Total: __/9	**Total: __/9**

Model: [redacted]

Correct Attributes	Incorrect Attributes
ADJ ○ Successfully cleared ADJ. ○ Unsupported enamel addressed. **Pulpal Floor** ○ Pulpal Floor Caries addressed prematurely. **Form** ○ Does not follow fissure pattern. ○ Overprepared in all directions ○ Good/acceptable angulation. **Overall** ○ Suggests stage is complete. ○ Identifies as not well prepared.	**ADJ** ○ Believes caries is still present at ADJ (discount small dots as not significant) ○ Erroneously identifies unsupported enamel. **Pulpal Floor** ○ Suggests that pulpal floor caries was appropriately removed. **Form** ○ Suggests fissure pattern was followed. ○ Suggests that angulation is poor. ○ Suggests an appropriate amount of tissue removed. **Overall** ○ Suggests stage is complete. ○ Suggests well prepared for stage.
Total: __/8	**Total: __/8**

Ranking Activity

Best 1	2	3	4	Worst 5

Appendix I

Comparative Study Procedural Task Rubric

This appendix shows the rubric used to assess performance on the procedural test described in Chapter 8.

Participant ID ____

Procedural Task

	Fail <40	Boarderline 41-50	Acceptable 51-60	Good 61-70	Excellent >71
Selection of Appropriate Bur	Does not recognise that different burs are available	Implies that different burs are used but gives weak or incorrect reasoning for their selection	Recognises that different burs are used but does not link this to the step being undertaken or explain their rationale.	Suggests appropriate bur size/design for each stage and the corresponding handpiece but doesn't suggest that different sizes may be appropriate during the procedure.	Recommends appropriate bur at each stage listing valid criteria applied in their selection. Describes risks of inappropriate selections.
Use of High and Low speed handpieces in caries removal/cavity preparation	Does not distinguish between high and low speed handpieces	Remembers high and low speed handpieces are available but is vague and/or inaccurate as to their use or role	Differentiates between high and low speed handpieces and describes their different roles in caries removal and cavity preparation.	Able to describe the appropriate role for high and low speed handpieces when removing caries/preparing a cavity and thei alternating and complementary usage.	Explains distinct roles of high and low speed handpiece in caries removal and cavity preparation. Describes nature of motions to remove tissue, appropriate cutting surface of bur and/or other considerations for safe handpiece operation.
Removal of accessible caries at ADJ	Does not recognise caries at ADJ as a discrete region to be addressed	Suggests that ADJ should be cleared by removing the caries but description of how this is achieved is vague, inaccurate and/or incomplete.	Recognises that ADJ should be cleared by removing caries and can describe the approach for doing so in broad terms	Describes how caries at the ADJ is addressed, suggesting appropriate instrumentation and providing some additional detail relating to depth or angulation.	Explains strategies for ensuring caries at ADJ is properly addressed. Describes importance of depth and angulation and suggests ways that these can be approximately monitored. Able to describe what `accessible' means in this context and when caries is no-longer accessible.
Distinguishing between infected/affected tissue	Does not demonstrate an awareness of what caries is. Regards caries as a homogenous.	Suggests a distinction between infected and affected tissue but does not describe how this relevant.	Recognises existence of infected and affected tissue and able to describe their relevance in caries removal.	Differentiates infected and affected tissue and is able to describe visual/tactile cues for distinguishing between them. Recognises when and where it is appropriate to retain affected tissue.	Explains infected and affected tissue and relates to corresponding to level of demineralisation and transition between them. Describes multiple ways of discriminating between infected/affected. Explains when and where carious tissue is appropriate to/should not be retained.
Management of unsupported enamel / Regaining access to ADJ	Does not demonstrate an awareness of unsupported enamel or that access to the ADJ is an important concept in caries removal.	Aware of unsupported enamel but does not suggest how it is relevant to the procedure	Unsupported enamel is recognised as an undesirable attribute. Consideration is limited to management and removal at the end of the procedure.	Describes that unsupported enamel is created when clearing the ADJ and that this impairs access/visibility. Explains how access must be regained throughout the procedure.	Explains how unsupported enamel is created when clearing the ADJ, the consequences of this and the steps to regain access to the ADJ. Describes role of dental probe in identifying unsupported enamel Recommends appropriate instrumentation and usage/motions when removing unsupported enamel.
Evaluation/inspection of ADJ and iterative nature of task	Does not suggest that ADJ should be inspected to ensure that it is cleared. Does not recognise the iterative nature of clearing the ADJ.	Aware that ADJ should be clear but does not describe an appropriate way of achieving this or determining if it has been achieved.	Recognises that ADJ should be inspected to ensure that it has been cleared. Describes an iterative approach to achieve this in broad terms.	Suggests appropriate timings of when to evaluate ADJ and describes its relevance to decision making when establishing a caries free margin at the ADJ.	Explains appropriate methodical approach to establishing a caries free margin at the ADJ. Describes relevant visual/tactile cues when evaluating the ADJ including the appropriate use of a dental probe. Details the value of using an iterative approach and how this facilitates a conservative approach.
Removal of caries from Pulpal Floor	Does not distinguish between caries at ADJ and Pulpal Floor. Suggests indiscriminate removal of caries or premature addressing of pulpal floor.	Aware of a distinction between caries at pulpal floor and ADJ but does not suggest its relevance.	Recognises that caries at pulpal floor should be addressed separately to that at the ADJ. Or: Pulpal floor not mentioned and caries removal description is exclusively focussed on addressing the ADJ.	Describes removal of caries at pulpal floor as a discrete and sequential step that should only be attempted once a caries free margin at the ADJ has been established.	Explains importance of retaining pulpal floor caries at this stage. Lists risks of prematurely clearing pulpal floor.

Overall Score _____

Appendix J

Comparative Study Quantitative Results

This appendix presents the quantitative results analysed in Chapter 8. In all instances Group A represents the Experimental group and Group B represents the Treatment As Usual Group.

J.1 Retention Activity

Group	Participant ID	Diamond Successfully Cleared ADJ	Diamond: Unsupported Addressed	Diamond: Retained Floor	Diamond: Base	Diamond: Follows Fissure	Diamond: Angulation	Diamond: Used Probe	Diamond: Is Good	**Diamond: - Total Correct**
A	1	1				1		1		3
A	2	1	1			1		1		4
A	3	1				1			1	3
A	4	1						1		2
A	8	1						1	1	3
A	14	1	1					1	1	4
A	16	1	1					1		3
A	17	1							1	2
A	19	1						1		2
A	20							1	1	2
A	22	1						1	1	3
A	26				1					1
A	27								1	1
B	5	1				1			1	3
B	6	1		1		1			1	4
B	7	1				1			1	3
B	9							1	1	2
B	11	1	1					1	1	4
B	12	1		1				1	1	4
B	13	1	1						1	3
B	15	1								1
B	18	1	1					1	1	4
B	21	1		1				1	1	4
B	23	1			1					2
B	24	1	1	1				1		4
B	25	1		1				1	1	4

Group	Partici pant ID	Diamond: Caries at ADJ	Diamond: Unsupported Misidentified	Diamond: Clear Floor	Diamond: Uneven Floor	Diamond: Not following fissures	Diamond: Poor Angulation	Diamond: No Probe	Diamond: Suggests Improveme nt	**Diamond: - Total Incorrect**
A	1								1	1
A	2									0
A	3							1		1
A	4		1							1
A	8									0
A	14									0
A	16								1	1
A	17							1		1
A	19									0
A	20									0
A	22								1	1
A	26			1				1	1	3
A	27							1		1
B	5							1		1
B	6							1		1
B	7			1				1		2
B	9					1				1
B	11									0
B	12		1				1			2
B	13							1		1
B	15									0
B	18									0
B	21									0
B	23							1	1	2
B	24									0
B	25		1							1

Group	Participant ID	Triangle: Caries at ADJ	Triangle: Unsupported Detected	Triangle: Retained Floor	Triangle: Base is smooth/even	Triangle: Follows Fissure	Triangle: Good Angulation	Triangle: Used Probe	Triangle: Recognises incomplete	**Triangle - Total Correct**
A	1		1					1		2
A	2		1	1				1		3
A	3									0
A	4	1	1					1		3
A	8							1		1
A	14	1						1		2
A	16		1					1		2
A	17	1							1	2
A	19	1	1					1		3
A	20	1	1					1		3
A	22		1					1	1	3
A	26				1	1		1	1	4
A	27									0
B	5							1		1
B	6					1			1	2
B	7		1						1	2
B	9	1	1					1		3
B	11	1	1					1		3
B	12		1					1	1	3
B	13		1							1
B	15			1						1
B	18									0
B	21		1					1		2
B	23				1					1
B	24		1			1		1		3
B	25	1	1					1	1	4

Group	Participant ID	Triangle: Believes ADJ Clear	Triangle: Believes No Unsupported	Triangle: Suggests floor should be cleared	Triangle: Suggests Uneven Floor	Triangle: Suggests fissures not followed	Triangle: Suggests poor Angulation	Triangle: No Probe	Triangle: Suggests Stage is complete	**Triangle - Total Incorrect**
A	1	1								1
A	2	1								1
A	3	1						1		2
A	4							1		1
A	8	1							1	2
A	14								1	1
A	16									0
A	17							1		1
A	19									0
A	20									0
A	22									0
A	26			1						1
A	27	1		1				1		3
B	5	1							1	2
B	6	1		1				1		3
B	7	1		1				1		3
B	9									0
B	11									0
B	12									0
B	13	1					1	1		3
B	15	1						1	1	3
B	18	1	1					1		3
B	21	1								1
B	23	1						1		2
B	24	1								1
B	25									0

Group	Participant ID	Square: Caries at ADJ	Square: Recognises/ queries unsupported enamel	Square: Retained Floor	Square: Base is rough uneven	Square: Follows Fissure	Square: Poor Angulation	Square: Used Probe	Square: Recognises incomplete	**Square - Total Correct**
A	1	1	1					1	1	4
A	2	1	1							2
A	3	1				1				2
A	4	1	1	1			1	1	1	6
A	8	1	1	1				1	1	5
A	14	1	1					1	1	4
A	16	1	1		1			1	1	5
A	17	1			1			1	1	4
A	19	1						1	1	3
A	20	1			1			1	1	4
A	22	1			1			1		3
A	26	1			1			1		3
A	27	1		1					1	3
B	5	1	1			1				3
B	6			1		1				2
B	7	1							1	2
B	9	1	1					1		3
B	11	1	1					1		3
B	12	1	1		1			1		4
B	13	1	1							2
B	15	1						1		2
B	18	1	1							2
B	21	1		1				1		3
B	23	1	1		1					3
B	24	1	1			1		1		4
B	25	1						1		2

Group	Participant ID	Square: Believes ADJ Clear	Square: Suggests floor should be cleared	Square: Suggests Floor is smooth	Square: Suggests fissures not followed	Square: Suggests poor Angulation	Square: No Probe	Square: Suggests Stage is complete	**Square - Total Incorrect**
A	1								0
A	2						1		1
A	3						1		1
A	4								0
A	8								0
A	14								0
A	16								0
A	17								0
A	19								0
A	20								0
A	22								0
A	26		1						1
A	27						1		1
B	5						1		1
B	6	1					1		2
B	7		1				1		2
B	9				1				1
B	11								0
B	12								0
B	13						1		1
B	15						1		1
B	18		1				1		2
B	21								0
B	23						1		1
B	24								0
B	25								0

Group	Participant ID	Circle: Caries at ADJ	Circle: Unsupported Detected	Circle: Pulpal floor addressed prematurely	Circle: Excessively cleared pulpal floor (no retained)	Circle: Does not follow fissure pattern	Circle: Good Angulation	Circle: Margin requires refinement	Circle: Used Probe	Circle: Recognises incomplete
A	1	1		1	1				1	1
A	2	1	1						1	
A	3	1								
A	4	1		1						
A	8	1	1					1	1	
A	14	1							1	1
A	16	1							1	1
A	17	1								
A	19	1							1	
A	20	1							1	
A	22	1		1					1	
A	26	1								
A	27	1		1	1					1
B	5	1	1							
B	6	1								1
B	7	1						1		1
B	9	1								1
B	11	1							1	
B	12	1	1						1	1
B	13	1		1				1		
B	15	1		1						
B	18	1		1						
B	21	1							1	1
B	23	1								1
B	24	1	1	1					1	1
B	25	1			1				1	

Group	Participant ID	**Circle - Total Correct**	Circle: Believes ADJ Clear	Circle: Believes No Unsupported	Circle: Suggests floor was appropriately cleared	Circle: Suggests appropriate amount of caries removed	Circle: Suggests fissure pattern followed	Circle: Suggests poor Angulation	Circle: Suggests margin well prepared	Circle: No Probe	Circle: Suggests Stage is complete	**Circle - Total Incorrect**
A	1	5										0
A	2	3										0
A	3	1								1		1
A	4	2				1				1		2
A	8	4			1							1
A	14	3										0
A	16	3		1								1
A	17	1								1		1
A	19	2										0
A	20	2		1								1
A	22	3										0
A	26	1			1							1
A	27	4								1		1
B	5	2					1		1	1		3
B	6	2							1	1		2
B	7	3			1	1				1		3
B	9	2		1								1
B	11	2		1								1
B	12	4										0
B	13	3								1		1
B	15	2								1		1
B	18	2					1			1		2
B	21	3		1								1
B	23	2							1	1		2
B	24	5										0
B	25	3		1								1

Group	Participant ID	Rectangle: Successfully Cleared ADJ	Rectangle: Unsupported Addressed	Rectangle: Pulpal floor addressed prematurely	Rectangle: Does not follow fissure pattern	Rectangle: Overprepared in all directions	Rectangle: Good Angulation	Rectangle: Recognises stage is complete	Rectangle: Recognises Poor Preparation	**Rectangle - Total Correct**
A	1				1	1			1	3
A	2	1	1	1						3
A	3	1				1			1	3
A	4	1	1	1						3
A	8	1	1		1	1				4
A	14	1	1	1		1		1		5
A	16	1			1	1		1	1	5
A	17	1								1
A	19	1	1							2
A	20	1				1				2
A	22	1				1			1	3
A	26	1	1							2
A	27	1				1				2
B	5	1				1				2
B	6	1				1				2
B	7	1				1				2
B	9	1	1			1				3
B	11	1	1							2
B	12	1	1					1		3
B	13	1	1			1				3
B	15	1								1
B	18	1		1		1				3
B	21	1								1
B	23		1			1				2
B	24	1		1		1			1	4
B	25	1	1		1	1			1	5

Group	Participant ID	Rectangle: Believes Caries still present at ADJ	Rectangle:R Believes Unsupported Present	Rectangle: Suggests floor was appropriately cleared	Rectangle: Suggests fissure pattern followed	Rectangle: Suggests angulation is poor	Rectangle: Suggests appropriate amount of tissue removed	Rectangle: Suggests more removal required	Rectangle: Suggests is well prepared	**Rectangle - Total Incorrect**
A	1									0
A	2									0
A	3									0
A	4								1	1
A	8			1						1
A	14									0
A	16									0
A	17						1			1
A	19								1	1
A	20									0
A	22									0
A	26			1	1		1		1	4
A	27									0
B	5								1	1
B	6									0
B	7			1						1
B	9									0
B	11									0
B	12									0
B	13								1	1
B	15								1	1
B	18								1	1
B	21								1	1
B	23								1	1
B	24									0
B	25									0

Group	Participant ID	Total Correct	Total Incorrec	Retained
A	1	17	2	19
A	2	15	2	17
A	3	9	5	14
A	4	16	5	21
A	8	17	4	21
A	14	18	1	19
A	16	18	2	20
A	17	10	4	14
A	19	12	1	13
A	20	13	1	14
A	22	15	1	16
A	26	11	10	21
A	27	10	6	16
B	5	11	8	19
B	6	12	8	20
B	7	12	11	23
B	9	13	3	16
B	11	14	1	15
B	12	18	2	20
B	13	12	7	19
B	15	7	6	13
B	18	11	8	19
B	21	13	3	16
B	23	10	8	18
B	24	20	1	21
B	25	18	2	20

J.2 Ranking Activity

Ranking Activity Data

Group A

	Participant 1	Participant 2	Participant 3	Participant 4	Participant 8	Participant 14	Participant 16	Participant 17	Participant 19	Participant 20	Participant 22	Participant 26	Participant 27
Diamond	1	1	1	1	1	1	5	1	5	1	1	5	1
Triangle	2	5	2	5	2	2	1	5	1	4	3	4	5
Square	3	4	3	2	5	3	2	4	2	2	4	3	2
Circle	5	2	4	3	3	5	4	2	3	5	2	2	3
Rectangle	4	3	5	4	4	4	3	3	4	3	5	1	4

Group B

	Participant 5	Participant 6	Participant 7	Participant 9	Participant 11	Participant 12	Participant 13	Participant 15	Participant 18	Participant 21	Participant 23	Participant 24	Participant 25
Diamond	4	3	1	1	1	1	5	5	1	1	5	1	1
Triangle	1	2	5	2	4	5	1	1	2	2	1	2	2
Square	2	1	2	5	5	2	4	2	3	5	2	3	3
Circle	3	5	3	4	3	4	2	3	4	3	3	5	4
Rectangle	5	4	4	3	2	3	3	4	5	4	4	4	5

J.3 Procedural Activity

Group	Participant ID	Selection of Bur	High vs Low speed	Accessible at ADJ	Infected vs affected	Unsupported and access to ADJ	Evaluation and Iteration	Pulpal Floor	Averaged Score	Overall Score
A	1	50	70	70	60	70	60	55	60	65
A	2	60	70	65	60	65	70	55	65	65
A	3	50	50	50	40	55	55	55	50	50
A	4	50	60	60	40	55	55	55	55	55
A	8	55	70	60	65	55	65	55	60	60
A	14	55	60	65	50	60	60	65	60	60
A	16	60	70	70	65	70	75	55	65	70
A	17	45	50	50	40	55	55	55	50	50
A	19	60	55	65	40	75	65	65	60	60
A	20	60	70	70	40	75	70	55	65	70
A	22	60	60	55	60	60	60	55	60	60
A	26	45	50	50	45	45	50	40	45	45
A	27	55	75	70	55	75	70	65	65	70
B	5	45	45	55	50	60	50	50	50	50
B	6	60	65	70	50	70	70	65	65	65
B	7	40	45	50	40	45	55	55	45	45
B	9	60	70	65	40	70	65	65	60	65
B	11	55	70	70	60	70	65	65	65	70
B	12	44	70	65	45	70	65	55	60	60
B	13	65	70	60	55	65	65	65	65	65
B	15	65	70	70	60	60	60	55	65	65
B	18	60	65	70	60	70	65	55	65	65
B	21	65	65	60	65	75	70	65	65	65
B	23	45	45	45	50	60	60	55	50	50
B	24	70	70	70	65	70	70	65	70	75
B	25	60	70	70	65	75	65	55	65	70

Appendix K

Comparative Study Interview Transcripts and Coding

This appendix provides edited participant transcriptions from the semi-structured interviews exploring participant perception of the experimental approach detailed in Chapter 8. Initial taggings are included which formed the basis of the extracted themes.

Participant Number	Transcript	Coding
1EXP	I thought these were really good [decon] worked through these, I liked them more, but then I think you could progress on to the full arch and that would be better. [acknowledges that decon not realistic] Yes, that's a very important thing [realistic] but I liked that [decon] more because I think this teaches you the core skill first and then you can practise it on the real thing. I think the realistic bit is still important but I think the arch just gets a bit busy, so this is [decon] just that you start off simple Realism: not sure what the realism gives - doesn't need to be but you're going to be ultimately doing it so it needs an element of realism Thought exercise 1 was a bit hard, because there was caries at the ADJ and I wasn't sure how to clear that. <exercise was to just show where ADJ was> Yes but I was still interested how to clear it. And I liked this one (exercise 3) thought this was very helpful in being cautious and just slowly working your way around [to clear the ADJ] Popups: I'm not bothered - don't mind waiting for people - not sure its realistic or if the graphics card could handle that. I think if you explain it beforehand it's relatively easy to understand. Prefer the tutor explaining it - I'd rather have a presentation and talk through it rather than the simulator - I don't think we should rely on technology to teach. I just think people teaching other people is a better connection and you get a better understanding. I don't mind waiting (for others to catch up) you can just reflect on what you are doing. <Prepare for test?> Yes. If I had done - obviously forgotten a bit - but I know more about the process. Perhaps full arch prepared better but more as a follow on. I just think the arch gets a bit confusing - I'm still happy with the training.	Prefers deconstructed Wants both Progression: Deconstructed First Management of cognitive load Realism: Differences Struggled with content Values ex3 Don't mind waiting Simulator performance Values tutor Don't mind waiting Prepared for test Management of cognitive load

	Just want the full arch to follow on. The bright colours are very helpful to work through it, Other: When started working on the infected and affected and stuff, you had a little speech thing whilst we were working. I don't think that was the most effective way to teach because I was focussing on that and listening as well, like, there's stuff you definitely said but I don't think I'd be able to say it again. I wanted to know the information but I was still working.	Values bright colours Simultaneous audio chanel: Difficulties
2EXP	I thought it was very helpful. The video with the showing you the technique was useful because you could compare what you saw on the video to what you saw on the simulator. I think my favourite was exercise 3 because it helped me visualise the kind of repetitive switching between the high and low speed the most and at the end exercise 4 helped consolidate that a bit more with also speaking with peers as well. I think this style is better to begin with because you get to see the different levels and then perhaps moving on to the exercises we did at the start of the year [in the taught course] once you have consolidated this knowledge, I think that would be more beneficial because you get a sense of the whole scheme of things put together. I think it [deconstructed] would be a good way of learning alongside our lectures and having both at the same time, i don't know if that would be possible, but I think it is a good way of learning instead of in a lecture in a way because you get to see it straight off rather than being taught the theory and then being exposed to this and trying to apply it (in reference to tooth structure and tooth morphology) I think doing this along side it might be quite beneficial (ie. to learn morphology and caries). If we were to do this alongside our histology sessions I think it would make a lot more sense and the knowledge you were gaining through them you could apply straight away - this has helped understand histology. Wouldn't say one is more important than the other, but once you've had a few introductory sessions the whole tooth is better to look at rather than looking at things in a more withdrawn perspective Preparation for test: Yes - especially exercise 4 Problems - bit laggy. Would have liked to have spent more time on them but we had time constraints.	Valued video intro Values ex3 Values ex4 Progression: Deconstructed First Wants both Decon supports learning Understanding Realism: Differences Prepared for test Simulator

	Another 10 minutes on each exercise would have been nice so you felt like you'd got to the end of it rather than moving straight on to the next. Other issues: The power point that accompanied the session was very good - just the time really. Popups: would be good to get feedback on what you have removed but it is also nice to just to be able to ask questions if you have any doubts but as an idea I think its quite good. Narrated: I felt like I was getting behind on what I was cutting through so you were describing what was ahead of me whilst I was still drilling the last step was trying to listen in whilst I was also doing it so it is quite tricky - can't keep up kinda thing. But it was good to hear a bit more information about what we were drilling through but maybe prior to doing it or just after. Exploratory style is good - very different - to being given a great load of information and then being told to do it. Can reflect on what you're doing as you do it. 1st exercise having the different angles of the ADJ was quite interesting to look at, not something I had really thought of as much in detail of how that the ADJ might look different depending on where the caries is in the tooth. Because in the previous exercises they were positioned very occlusally and didn't really spread very far out so maybe more exercises like that.	performance More time Valued video intro Open to popups Simultaneous audio chanel: Difficulties Understanding Unconsidered aspects Understanding Values ex1
3EXP	I feel like these exercises really help highlight the ADJ even like the colour coding and stuff I think its more helpful. I feel like with the ones that we were doing earlier that it brushes past the ADJ but with this it kind of highlights the importance. I feel like both would be good, a combination of both would be good, because when you drill a tooth in real life you're not going to see blue, you're not going to see these like different colours so its not realistic in that sense but I think if there is a lesson on the ADJ then I think a combination of both… yeah. Learn it on this then apply it. I especially liked this exercise [exercise 4] it was good that it was at the end because it prepared us for knowing what was good or bad. The exercises prepared me for the test - especially the last question because I feel like if we didn't do the last exercise i might have struggled.	Values bright colours Unconsidered aspects Wants both Realism: Transfer Concerns about transfer Progression: Deconstructed First Values ex4 Prepared for test

	<Following on from that - being shown a tooth and being talked through the clearing process would prepare you less?> Yes - that's what I'm trying to say. Popups: I feel like I like to learn from experience, if we were told to do it with the pop ups then do it again without them - they wouldn't bug me - but I want to explore it myself. 2nd time you do it you get the pop up - so it's like a guide so you can see if you've taken in what the guide is saying. I think this was genuinely the best teaching approach because we were doing one tooth at a time but in this one (ex 4) we can even like compare them. Last comment is that this is really effective.	Prefers deconstructed Timing of popups Open to popups Prefers deconstructed
4EXP	I think this [decon] is probably better at first because you can actually visualise what you are doing because I remember when we first did the first taught sessions, I mean it wasn't bad but you didn't *really* know the right order to do it. I remember when I did it I just drilled straight down and you might hit the pulp. I didn't really realise you were meant to do those movements so I think this is the probably good at first but obviously you can't continue doing that when you get more advanced because that's not what teeth look like. I think once you've got this foundation down, once you understand what you need to do, thats when you could do like we did in the taught sessions. I think focussing on the concepts because honestly its quite hard to do it realistically and i feel it would probably be better to do it on plastic teeth. I think the activities prepared me for the test. I really liked this one [exercise 4] because its quite interesting to see, because it's interesting to be able to compare different ways that it went wrong or different ways that it went well. And I liked this one [exercise 2] because you can see how far it went and its got all the different colours to distinguish between the infected and affected. I like the idea of having pop up boxes because everyone worked at different paces and I remember being done and just sat there wondering what to do so maybe if there were pop up boxes but maybe if you could turn them off so the person can choose if they had them maybe that would be helpful.	Progression: Deconstructed First Prefers deconstructed Understanding Progression: Deconstructed First Understanding VR vs phantoms Values ex4 Values ex2 Values bright colours Open to popups Timing of popups

5TAU	I think as an approach it is very systematic and to the point which is quite good to get an understanding of the basics. In terms of improving the approach people will have different shapes so seeing if there are any ways that the technique could be adapted for different types of cavities but otherwise I did like it because it was very systematic and to the point. My only obstacle was trying to use the actual thing [makes an operating handpiece gesture in the air] because I find it quite difficult because the shape of it and the weight of it is not like a normal typical dental instrument so it's quite limiting in terms of the real life applications.	Valued TAU Simulator performance
	I think the goal isn't necessarily to replicate the actual clinical environment but the habits you will pick up on that [the simulator] if you then somehow use those in actual clinics then that might be an issue. For example I can't do a finger rest on that at all so i'm used to not doing a finger rest but sometimes I go to the clinic and I cant do a finger rest either so that might be just me.	Concerns about transfer Realism: Differences
	If you move the actual instrument a centimetre it moves a mile on the screen sometimes, thats what it feels like. I think its just the size of the tooth is so different to what you will have in real life and you're going to be used to working on a larger scale so its' just understanding that when it comes to real life clinical practice it will not be completely applicable but it is good to understand the basics and the steps.	Simulator performance Concerns about transfer
	Shown decon: I think this is better because you can appreciate the structure of the tooth and relate that to the steps and why you needed to do a certain step like removing unsupported enamel and things like that. When you are doing things systematically you're not able to appreciate it properly because you don't have the actual view but here you can see the actual cross section of the tooth and can appreciate where the unsupported enamel is and where different structures are. And when it comes to you actually chipping away at different parts of the tooth you know where you are now and it will make more sense as well.	Prefers deconstructed Understanding
	I think its better for you to understand the basics through this first before you move on to something like that because at least this way you understand the fundamentals of the different parts of the tooth, the structure and then you can move on to the whole tooth itself once you've learned everything. I would do this but I would still do the whole tooth afterwards.	Progression: Deconstructed First Wants both
	Intervention prepared me but i have a very bad memory but yes it did definitely.	Prepared for test

	I quite enjoyed it because I needed the extra practice anyway but I did really enjoy it though.	Enjoyment
6TAU	I think maybe one thing could be better, maybe a video demo, it doesn't have to be perfect but maybe a video talking though how my hand should be gestured or just watching the process of a professional removing it because in the session we were just told what to do you know like keeping the hand straight and the finger rest or something but maybe a demo could help but yeah because I think this kind of thing you need to gain experience on it and hands on experience so I think this session was a good opportunity for me to practise and I think i have improved a bit at using the vr so maybe just it's just a really good opportunity. I think a very good thing about this one [exercise 3] is the colour coding because the exercise we did the enamel was all white so we had to do it little by little so we had to figure it out and do and slowly work out when the ADJ is cleared but this one maybe that could be good one to have a go at clearing the ADJ - making a caries free margin then the 2nd try could be the one we did - no longer having the colour coding as a guidance yeah I think this one the colour coding is a real help. I think to be fair in real life there is no colour coding so that could be integrated but maybe only for the first few steps because you still have to move on to "no hints" like this so I'm not sure. [emphatically] yes I would appreciate something like this being part of the course Trade-off: Difficult question - a mixture of deconstructed. Ideally want 1, 2 and 3 [points at exercises on print out] - two might be enough still want to have more time to practise on realistic ones Want realistic because it's realistic, the session we had we jumped straight into realistic it's just it's important to feel confident without hints like this so a couple of exercises to introduce like transitional to the harder ones would be good Exercise helped prepared, it was very clear, all the techniques and things to be careful of were straight forward and clear to understand its just having the confidence to do them myself. Other: It's all been really good its just we normally only have 2 VR sessions so maybe just one more because	Wanted demonstration Hands on Values bright colours Values ex3 Understanding Progression: Difficulty curve Realism:Transfer Progression: Deconstructed First Wants both Concerns about transfer Wants both Prepared for test Enjoyment

	I really enjoy it.	
7TAU	I think maybe like potentially going through it going through the steps you do then sending us away to do it and keeping an eye on us then and then again going through what we did right and what we did wrong and then giving us another go I think that could be helpful I can't remember if you did that or not. I think maybe having a go on the same tooth so we can feel more confident with one area of the mouth would be helpful. To be honest, I feel like it doesn't need to be as realistic as possible its more about the techniques we are using because if we get into good habits now it will be useful but I also feel that using the simulators can be a bit of a, it throws people off because like if we have a lot more sessions on it we will like get used to it and then we are actually transitioning to actual patients to actual teeth like it feels a lot more different and we just might get used to the feedback we are getting when we move these instruments and then we have to get used to something else. So I feel, yeah, I'm not too sure, I do like the idea of VR especially in 1st year when we don't get to do a lot of clinical stuff I just feel that it should be made apparent to people that it isn't like the real thing. Decon: I feel like it could be something that you could use that you do this session 1st then you complement that with our session, just for completeness but I do like how this is a step by step approach and it actually explains to us what bits are caries and which bits aren't; which bits do we use a high speed which bits do we use a slow speed so we can get used to it but I do think that after this I would also be feeling a bit left a bit confused as to like what would I do on an actual tooth so like having this session then going back and doing our session and drilling on actual teeth that aren't broken up or anything I think could be useful, yeah I'm quite a visual learner and I like doing things whilst I'm learning so I think this does definitely appeal to me. I think you need the full experience to have the biggest impact. [Prepared for today?] Yes, for sure	Wanted demonstration Realism:Transfer Concerns about transfer Realism: Differences Progression: Deconstructed First Wants both Prepared for test
8EXP	I think this session was more well executed because we were taught what needs to be removed and how it looks like and the technique behind it and it introduced	Session delivery Understanding

	us to some things we haven't learned before so I find it better in the sense that it let us have a proper [?] rather than just going in and not knowing what we were doing. The exercises were good because they were in stages so in terms of difficulty the difficulty increased which is better instead of just starting with a really hard exercise. [negative to this style] no I don't think so, the only downside I can think of is students who don't get to try this exercise may feel like they don't have the same experience they may feel like they are behind. I think this is better because you can see what happens after each stage and you can see the cross sections so I think it helps us visualise and understand the anatomy of the tooth rather than just doing it on a model. [Prepared?] Yes definitely because I think I wouldn't have known where the caries are ar the adj or how much to remove so it was definitely helpful. [Popups] I think it would be useful because it gives a student a gage of progress so instead of just drilling and not knowing where to stop. I don't think the prompt being interrupting is an issue because it would let the student know how they have done. A 1 to 1 discussion could be good with the tutor so that the tutor can provide more feedback. I think it would be really useful for potential first years.	Prefers deconstructed Progression: Difficulty curve Prefers decon Prefers deconstructed Understanding Prepared for test Open to popups Values tutor
9TAU	I think it was quite effective I think it was kindof nice having the presentation alongside it with images of before and after and what you were looking for. Again having the more distinct colours was useful for beginners even though I know it's not like that in real life but it does help in the beginning. Maybe one thing that might help but I know that its probably time conscious is maybe seeing it being done first on the VR machines and then doing it ourselves but I know it might be difficult in terms of everyone watching the angles of the machines and things. But apart from that I think it was quite beneficial and I did enjoy it. It was a lot more fun than being sat watching a presentation being kindof info-dumped on how to do it. [Realistic?] Yeah, I think maybe at the beginning as we are fresh with no previous knowledge I think having the more helps you with understanding with more differentiation between colours and things and then	Valued TAU Wanted demonstration Hands on Understanding Values bright

	kindof slowly progress to more realistic the more practice we do. [Prepared?] Yes definitely, I think so [Decon] Potentially I think from a starters POV especially this one and this one [ex 2 & 3] just getting an idea of clearing the ADJ and maybe having it as a cross section might have been an easier starting point rather than a whole tooth. So potentially as a start but then, as I said, the whole tooth. [Anything else?] It was all kindof well explained - I think it's just the temperamental-ness of the machines but I don't think that's anything to do with how we are taught and any other understanding so yeah. I'd definitely do it again.	colours Prepared for test Progression: Deconstructed First Wants both Simulator performance Enjoyment
11TAU	It was all pretty clear to me and the steps made sense so it linked why you did the different steps so it was like easier for me to remember. Personally for me I'm the type of person that has to understand why I am doing the step and it sticks in my head easier so I think that was one thing that was good about that lesson. I compared to the one I've done before it's like more hand holding you throughout the entire process I guess I feel like the way that I did it before was like you need to know a bit more information before you go into the session because like [gestures at ex3] even if you don't know what you are looking at it is all labelled clearly for you and there's different sections and it's labelled ADJ I mean that for intro lesson I think this is easier for people to see it and I know what I'm looking at compared to [?] what are all these different parts of the tooth? And I think this that the one I did before is like it simulates the real world more in that nothing is labelled for you and you have to feel your way through it and learn as you go but this is more like we label everything you need to know so there's pros and cons I guess so maybe you need to mix both of them together in the end. [Acceptability of decon] I guess this helps you visualise in a clearer way because like for example when you are learning how caries spreads through the teeth and things like that there's no real visual representation we just feel through the dark and we're just like oh it spreads this way. I think in terms of helping me visualise and understand if I'm doing it right I think this offers more reassurance that yeah I'm following the procedures properly	Understanding Hand holding Understanding Values bright colours Wants both Management of cognitive load Unconsidered aspects Understanding

	because at first when I was starting the 1st exercise I was a bit in the dark, what am I doing, am I doing this right so I think this clears up that a bit. [Prepared?] Yes - they achieved their goal	Prepared for test
12TAU	I think it worked well it's just like the simulator was a bit tricky to use compared to use just because like compared to the previous sessions I was using the high speed at one point and it wasn't taking any enamel off and then all of a sudden it started working and just went like crazy. But teaching wise I really liked that we got the walk through then you got to put it into practice bit by bit then you get to put it all together at the end on your own. Which was helpful. [decon] I feel like I kinda prefer this… just because like it's kinda like better with the ability, kinda like having the cross sections and then obviously like when you get used to knowing what you need to remove and everything then you can kinda move on to like the whole tooth potentially. It's because we did spend a lot of time trying to get the angles right to see it and stuff. I think it would be good to have a more basic foundation like this and then move onto the whole tooth. It doesn't have to be a slow progression it could be faster. I don't think its important to be realistic when you're starting out but it should be realistic later. I guess it gets you put into the same mindset, I guess it's easier to apply when you see it on a phantom head or a real patient because its the same. Especially if it's in like a mouth because you can't turn it all the way round to the back of the head and start working there [referring to ability to rotate the view to achieve an impossible angle]. Prepared for today? Yes I would say so. [Other points - Not applicable to this study - were in reference to the scheduling of taught sessions in 1st year]	Simulator performance Valued video intro Prefers deconstructed Understanding Progression: Deconstructed First Progression: Difficulty curve Realism: Differences Concerns about transfer Prepared for test
13TAU	Yes I think overall it was quite helpful. I liked going through the powerpoint at the start and if I remember correctly there were quite a few images on there so I think that was quite good and then just doing it step by step; starting with the high speed and then waiting until everyone had finished with that and then moving on to the low speed and just doing it in a process	Valued video intro Hand holding

	rather than getting the steps and then doing it all at once I think that was quite helpful. Different speed students - I do think that is the downside of it but I don't know how you would change that system. Yeah. [Prepared?] Yeah definitely. Without what we'd done I wouldn't have a clue which one of them was the best or the worst so it did help with the understanding of what to look for and what the good things and the bad things to do are so try not to make it really slanted and that. Quite different, I think this one actually looks really good I think I like that you have the tooth divided so that you can see inside especially when you can see that you are reading *ADJ* as letters. I think it makes it quite obvious. I think the one we did is probably more similar to when you are actually having to do it because it is the whole tooth but I think as a start maybe this one to just get the steps in and see it then maybe progress to the one we did afterwards. [Want both?] yes… yeah I do. I think the one we did prepared us for the task because that looks more similar to what we could see on screen so you get its more similar in that sort of aspect. [Should be realistic?] I think maybe at the start not so much, maybe at the start focus on the different aspects of it and seeing it from the different views but I think as you get later on I think it is important to see it as realistic just that you are prepared when you get an actual tooth it doesn't look really foreign to you. No other comments, I think it was a good task to do.	Prepared for test Values deconstructed Values ex1 Understanding Progression: Deconstructed First Wants both Prepared for test Realism: Differences Progression: Deconstructed First Concerns about transfer Enjoyment
14EXP	I would say that I like it broken down, the second method [decon] is better because its then aiding you in what to do because it is hard using the virtual reality machines because it's obviously not the real thing. I think it's better to do baby steps with this rather than going to the patients mouth because trying to orient the patient's mouth was hard but doing it when the teeth were lined up is kinda hard as well. And I kindof think we should be taught a better plane to use it in because some people were like using it not tilted but you try and get it perpendicular to the tooth but then	Prefers deconstructed Realism: Transfer Simulator performance

	like the tooth could be like this and I could be like this [gestures angulation with probe against tooth model] but in reality there might be a better way for me to do it so if I knew I'd just do it. I would like concepts first and then into the real thing because it would be easier to pick up. [prepared?] Yes I'd say so it made it much more easier to recognise and stuff. I think all of it was effective. I think the middle on [ex 3] is a bit hard because it's like pro-style but I think it helps a lot because when I was thinking back today I was remembering that oh like you need a slow hand piece as well and the fast hand piece. [ex3] brought this memory back because I remembered you need the fast handpiece to get rid of the unsupported enamel I'd say it helped a lot really. I think an emphasis on the circular motions in the presentations, I don't think people do that and just do their own thing but it does help when you do the circular motions. I think just keep emphasising it between each slide yeah slow circular motions. [Popups] I think you can have it but give people the choice of if they want that option or not. I mean its not bad having it there but if one person doesn't want it they can turn it off. I'd say in summary I had a fun time, I learned what the ADJ was I guess I should know that but I guess I learned something new and I'm excited for 2nd year it was like more practice as well.	Progression: Deconstructed First Wants both Prepared for test Understanding Values ex3 Attentional focus Timing of popups Understanding Enjoyment
15TAU	I think it was good because that's what we're going to be dealing with. I think I found it, I think its because I hadn't remembered from the previous sessions, what the difference between the healthy dentine and carious dentine like the colour difference. I didn't find that obvious at first like what I should be doing. I don't know if more obvious colours just for this exercise would make a difference or not. [decon] I think so (it appeals) I think the colour coding makes it more obvious and I think having the cross sections is good for visualising the ADJ better and the idea of how the different levels and there's different points to be removed [gestures at ex 3's gradient points] I think that's good. [Realism] I think that it's pretty good because as long as that is paired with more traditional way of looking at	Realism: Differences Values bright colours Values bright colours Understanding Values ex3 Progression: Deconstructed First

	it, it doesn't seem like it's too different from what you would actually be doing and I think it's quite a confusing thing to understand so seeing it visualised in this way makes it a bit more obvious. [Deconstructed as 1st year content] I think it would still be fun and enjoyable and I think it would make it easier to learn about it. I don't know if maybe you did something like this first and then had a quick go on the normal way of doing it - then you know what you are doing. Then you can try it for real. I quite like the visual way of looking at it, I think it's easier to understand for me. [Popups] I think it is an interesting idea I don't really know what that would be like it might be quite good for progress, because there's lots of people in the class to have something so you know you're doing it the right way… hopefully… yes a bit of encouragement that you're making progress but I do think it's nice having a tutor giving feedback and stuff too. [Prepared?] Yes I think so, I don't know how it would compare to this I was a bit confused at the start but once I knew what I was doing. I found it hard to engage with the Powerpoint at the beginning, when I started I wasn't entirely sure what I was doing. But then when I realised the procedure for doing it it made more sense. Struggled to relate the powerpoint content to the task until the walk-through started.	Concerns about transfer Realism: Differences Progression: Deconstructed First Wants both Understanding Values tutor Open to popups Prepared for test Disliked video intro Understanding
16EXP	I think I preferred doing it like this, having everything broken down, because when you have the tooth in the full arch or like a patient there it's difficult to like not to like in a VR context it's difficult to not put your hand through the patient's cheek like you don't have that same tactile feedback. I guess that's only something you can do on phantom heads where you're able to feel it whereas doing it separately you are able to focus on the technique and exactly what to look out for and breaking it down step by step so you can see the different colours, like more examples of how to do it, I felt like that was more.. well, for me it helped a lot more to break it down rather than jumping straight into it. I think having the full tooth at the end to maybe consolidate what we've learned on these separate teeth. Because this breaks it down step by step and for me I've got to, I'm quite slow, so it takes me a while	Prefers deconstructed Values bright colours Management of cognitive load Progression: Deconstructed First

	to pick up on there may be certain stuff so being able to do like each step and different examples before going to the full tooth it like helps, it helps a lot. [vs conventional] I personally prefer it broken down, I feel like this is part of the the only part of dentistry that we do in 1st year. I don't know anything about dentistry, but I don't think I'd feel short changed, I'd prefer to learn how to do it step by step and having it broken down and maybe having a separate session maybe afterwards where we consolidate what we've learned and do it on a full tooth. [Prepared?] Yes I think so I think maybe having a powerpoint afterwards because I feel like I've forgotten the key words so if I can read through it a few times it can help me consolidate what we did practically. It felt like the machine struggled and was a bit laggy on [ex1] sometime it was a bit frustrating but when you got in and were making small movements it was ok. I really liked the last one [ex4] we could see it and discuss what needed to be changed to consolidate what we did earlier. [Popups] I think it would be good, definitely in terms of getting on in our own time because sometimes I would be waiting or I'd be rushed because I was taking too long but with the pop up boxes it will probably pop up at the most annoying time, they usually do, but maybe a mix of you guiding us and maybe the pop ups for just key points rather than everything. Rather than every step it just being things to look out for, because you might have said it when explaining it to us but maybe it's something we'd missed. Just have to be careful with the amount but in general it's a good idea I'd be keen to have them. Other - NA - requested more SIMS sessions. I really enjoyed it	Progression: Difficulty curve Prefers deconstructed Wants both Prepared for test Simulator performance Values ex4 Open to popups Timing of popups Enjoyment
17EXP	I guess like seeing the individual tooth is easier but I found these exercises harder than the ones that are taught in our normal session - because we've not been taught how to prepare the ADJ before, we've not learned that before and we won't do that until next year so it's all a bit new. The concepts are the same as what we've been taught	Struggled with content

	and we're just applying that to these exercises but I think just understanding what it is we're doing, why we're doing it, what a properly established caries free margin looks like that was all a bit new the techniques were much the same. I think its' good to start on individual levels so you can fully appreciate what is going on on that individual tooth and get a good understanding and then the next step would be the full arch. I think there's a place for this kind of approach, because if you just went straight into doing it in the arch it might be a bit overwhelming because you don't know what you are doing because there's a lot of teeth and you can't see what you are doing properly so to see it break it down and just have one tooth is like baby steps. Prepared? Yes, especially this one [ex4] and when we thought it was green or red because there was one I wasn't sure about but when we were discussing it I was like yeah that makes sense now. [Improvements?] Maybe having another session before this one so everything we learned in the 1st session we can try and perfect it in the 2nd one. [Popups] Its hard because I think its very individual some people might find it very beneficial but some might find it quite annoying if it kept popping up. Personally, I quite like if I wanted the feedback so I'd press the button rather than it being automatic. I think I would like it… I prefer this approach because it was on a more individual level and you could just focus on that one tooth and see what you were doing because when it's in the arch that's more difficult so especially when you start I'd like this. I quite enjoyed it.	Understanding Progression: Deconstructed First Wants both Management of cognitive load Hand holding Values ex4 Prepared for test Understanding Understanding Timing of popups Prefers deconstructed Understanding Enjoyment
18TAU	I think VR is an effective way of introducing clinical skills before you go to a real patient its a very good thing that Sheffield has that however I think there could be a thing that a few changes in the way that they teach it. I think with that session especially with the video that you had I think it was clear like the steps you had to go through but I guess maybe in the sessions when you focus on particular task you can forget what the next step is or what you're meant to be doing here or there. So maybe broken down into smaller steps where you might explain that this is the	Values deconstructed

	procedure to use the high speed drill to get down through the enamel and then once your students have finished with that maybe they could then explain the next step so its more fresh in your mind so you're not having to think back for what is the next step because some of these procedures its not just so simple like you have to do two things some have multiple steps i feel for those sorts of procedures if you were to break it down into smaller steps students would get a better learning outcome from it. But in terms of the actual facilities, I think some of the machines are a bit lagging but other than that the teaching method is quite effective but maybe break it down into smaller steps there's lots of steps involved.	Management of cognitive load Understanding Simulator performance
	My method would be like the enamel, for example, if you were to explain what we were doing in that step once students have had a go at it. So the first step is to use the high speed right, so once you've explained how to do that you stop then then you explain what the next bit is so you're taking it one step at a time rather than everything all at once and telling them to have a crack at it. I think that's just how I'd prefer it in my experience.	Management of cognitive load
	<did you speak to anyone in the other group?> No not at all <explains other arm of study uses deconstructed> oh wow that's very interesting - no honestly I've not spoken to anyone.	
	[decon] In my opinion yes I think that would be a much more effective way of getting students to understand what the step by step procedure is of preparing a cavity. I think even though when you do one in real life there's no letters or colour coding on there its definitely a great way to get it into students heads that sort of memorisation down so that when they are doing it they'll have these things to refer to and I think it's much more memorable learning these steps that way when it comes down to actually doing it. So I think yeah, in my opinion it would be a much more effective to yeah sort of do it like this where you're breaking it down into steps and testing to see if they've learned anything from it. I would say I'd prefer that a lot.	Prefers deconstructed Understanding Attentional focus
	I don't see why it would have to be solely one for example if you were to do this exercise first and explain what everything is and have it so that the students have a really good understanding of what each particular step is and what each section is and then go on to the exercise that I did to confirm that they understood it properly and could do it without holding their hand the whole way through. I think that would be a much more effective way, I don't see why it	Progression: Deconstructed First Understanding Wants both

	has to be just one or the other. [Do this in future years?] I think if its 1st, 2nd and 3rd you then you're still learning everything you haven't done these things a million times like you would if you were a 5th year so say for example you're in 3rd year and you've done this procedure lots of times then doing it solely in the 1st method, because they don't need it, they don't need someone holding their hand the entire way but definitely for 1st and 2nd year where these procedures are brand new to them and they haven't done anything like this before then I feel its vital that they have it broken down to make it as easy as possible for them because it's' the understanding part that is key isn't it. [Prepared?] Yeah I'd say so, if I hadn't taken that session - I'd say yeah. I think the study was fun and it was a good opportunity because the VR suites aren't that available so the extra practice was great and I learned a lot more from that session so yea I thought the whole thing was good. The only thing I would say is that some of the machines need looking at.	Progression: Deconstructed First Hand holding Understanding Prepared for test Enjoyment Simulator performance
19EXP	I think it worked well overall in terms of letting the information to be able to attain it[?] because I guess like instead of just lecturing on or teaching about it having hands on with the VR itself there's something about it that makes it stick to your head better I guess. In terms of the specific exercises themselves I liked the ADJ part the way that its highlighted in different colours that helps for when you're beginning I guess to be able to like tell. The only thing I would say I don't know if it's like it's probably just like the clinical aspect in most of the tasks some parts were in the natural colour which is being used to denote or highlight a certain area I guess I would be slightly uneasy if I was looking at an actual tooth as opposed to one on the simulator because in the exercise some parts were like they were like the areas you had to work on were highlighted in a very obvious colour so I know that I should work on there because you'd see it and you'd know. But opposed to if you saw an actual like tooth because using the VR it highlights those areas bit in an actual tooth it wont so you'd need more experience to tell. The tasks themselves though I think worked well in terms of using the simulations probes and rotors and whatnot but I don't know if this is a thing with the machines because some work better than others which might affect the experience.	Hands on Values bright colours Concerns about transfer Simulator performance

	I guess if I could only pick one; I guess the only reason I would say realistic probably more is due to us not having much experience in the first place so the more opportunities we get experience the quicker we could identify it quicker on (a real pt). I'd say probably both, this probably helps first to establish the concepts[?] of what to do but then you could move on to the more realistic when you have confidence in what you are actually doing before you even go on to it. For the concepts itself this [decon] is better because it points it out to you more with the colours and hands on experience. I'd still say the realistic is important. Mainly because as a 1st year student we don't have much experience with realistic yet as good as being able to identify it on here would be we don't have much knowledge about how it will look realistically so we don't have too much of a great comparison and it creates anxiety of will you be able to realistically identify it and whatnot. [Prepared?] Oh yeah definitely. [popups] it seems like it would work i guess, the only thing would be that it can't answer any questions that you have but aside from that point I can't think of anything. [verbal drill along] worked well for me I'd say I didn't have too much of a struggle to pay attention	Realism: Differences Wants both Progression: Difficulty curve Values bright colours Hands on Realism: Transfer Concerns about transfer Prepared for test Values tutor Simultaneous audio chanel
20EXP	I'd say I found this really helpful [pointing at ex4] at the end to find which one was wrong and this [ex1] was helpful to see the ADJ and also that it's not always in the middle so that was helpful. This one [ex2] I remember you were reading from a script whilst we were going through, I was struggling to keep up a bit, and this one [ex3] was helpful to see how far back you need to go because I thought you just drill a hole and obviously you don't. I find this to begin so you can see how far, because if you had a full tooth you wouldn't be able to see the whole thing at once but as you get more used to it then to see a full tooth because I think this teaches you that you can't angle the bur and you have to go back like this but on a full tooth I'd be a bit like this [gestures off angle bur] You wouldn't be able to see all this all at once on a full tooth and I found that really helpful. I think that initially would help. I like this visual to help but it will be a full tooth.	Values ex4 Values ex1 Unconsidered aspects Values ex2 Values ex3 Simultaneous audio chanel:Difficulties Understanding Understanding Progression:Deconstructed First Unconsidered aspects

	[Why realistic?] Good practice before going onto patients getting like the feel for it because its brand new concepts. [Prepared?] Yeah. This consolidation at the end prepared us most. [popups] I think it sounds good because it's more individual it's not a group exercise and you'd be able to go at your own pace so maybe you'd get better feedback so if it's giving instructions I think it would need to come with pictures or a little simulation of how to do it. I think it sounds like a good idea. Every time I do it I seem to have a really slow simulator.	Progression: Difficulty curve Realism: Transfer Prepared for test Open to popups Simulator performance
21TAU	I really enjoyed it I really like using the virtual simulation suite its probably one of my favourite parts of the course so far to be honest you just feel like you're a proper dental student don't you. I've really enjoyed it so I really wanted to participate in this study. [Prepared?] Yes I definitely think so I think it was very informative, you went though everything very well and very clearly yes I think it was quite good. [decon] I think this does seem good and with the bright colours at least initially so you can see more to help you understand it more but I think I prefer the other way its just a bit more realistic and like what you're actually going to be seeing so I don't know if you could use both, I think this is good at the beginning but its more realistic with the other one, I know it's not real, but you're seeing the actual decay whereas this is bright colours and obviously its not going to be like that but I think either way is… I think this would be really good for developing your understanding at the beginning to be honest but so's the other one to be honest… I think they are both quite good but one is more realistic so I prefer that. [why realistic better] I just suppose when we actually qualify that's the sort of thing we are actually going to be seeing erm and also I suppose when you have a whole tooth that is what you are going to be seeing whereas this you have a section cut out and it's not going to be like that but I would like to try this way as well to develop my understanding. I think probably a bit of both - to develop your understanding because it shows the layers quite well	Enjoyment Prepared for test Session delivery Values bright colours Realism: Differences Prefers TAU Wants both Concerns about transfer Understanding Progression: Deconstructed First Realism: Differences Progression: Deconstructed First

	that would be really good so you can say that is here but you need the other way as well so you can see what its actually going to be like. No opinions that I've just enjoyed being part of the study.	Wants both Enjoyment
22EXP	It was good. Yeah it was fun to do. I feel like they were quite quick, like we didn't have much time to do each one. I think there was a bit of a rush because someone didn't turn up or something. But, yeah I feel like I learned a lot from it, like doing doing the different angles [gestures at ex1] like I hadn't thought about that [points at ADJ in the cutaway on the mesial aspect of the model] and I like the colours because I feel like it makes it a lot easier to see but then it's hard at the same time because when you have an actual tooth its not going to be that same colour. But then it's useful when it's more like that… I can't think of anything more to say… Preference? Its kind of easier to see like what you are doing like this especially first year because you don't really know that much about like how deep to do it and you can see better on this which I think is good. But equally seeing the whole tooth because when you're actually doing it you're not going to see it. So yes. Kindof both. I know when we did it [taught session] we had it in a mouth but that was kindof hard on the VR because you can't get, and I'm left handed and I think it was on the side like and I couldn't get to it. So I found that hard when it was inside the mouth. I think this is good, using it as a way to learn it and then actually do it on the phantom heads because sometimes it, not laggy, but not exactly the same as it is in real life so I think it's good to learn it and practise it. And like know the theory of it, doing the steps like this then you can put that theory into practice. [Prepared?] Think so yes and doing it in steps was good. [Popups] I feel like that would be good, I was like one of the slower people so I feel like if you're a bit behind you panic and feel like you need to go a bit faster and then you do it wrong and in the same way if people are really in front they are just going to be a bit bored so I think that would be good but you probably need obviously need someone coming along in case you're confused what it is saying and maybe demonstrating it but I think it would be good yeah. [Other thoughts?] (some comments about an issue	Enjoyment More time Values ex1 Unconsidered aspects Concerns about transfer Understanding Wants both Concerns about transfer Simulator performance Progression: Deconstructed First Realism:Transfer Prepared for test Open to popups Values tutor

	with the simulator in the timetabled session prior to the study) Can’t really think of anything else, it’s quite fun to do because if you mess up its not stressful and I think if we went straight into the phantom heads and started doing it I think it would be kindof stressful where this builds up nicely. <does this kind of exercise help?> yes. Its good because there’s lots of different things and if you mess up this… it’s good…	Enjoyment
23TAU	I guess maybe more of a demonstration I don’t know if that’s live or maybe just a video like visualising more what is happening I think it would be useful to see someone’s hand at the same time as watching them drill on the screen so you can see how the two link up. [Decon appealing?] Yeah I especially like with the colours and that like its labelled I think when you first start out you’re a bit confused about what you are seeing so I feel like that would be a little bit clearer because like I think it was a lot more easy to do the simple exercises that we did [in the taught session] like getting rid of the cross but when we moved to the tooth I think it was like it was definitely a bit of a jump so yea I think like because that still doesn’t seem like 100% real, you still have things that are there to help you so I do think that would help you. I guess I think if you start off like that [decon] I think its going to be easier to transition into the more realistic I wouldn’t like keep to that throughout the entire thing I’d definitely switch to it looking more like a tooth. But I think to start off it could potentially make it a bit easier. <just do whole teeth in the phantom lab or keep whole teeth in VR?> I think there’s still value in having a full tooth in VR for sure. Difficulties? I think I just generally have difficulties with VR I don’t know what I do but it always ends up going wrong. [Compare TAU to EXP approach - pros/cons?] Just going through things step by step and getting help at each stage knowing what you’ve done well or what you could have done better as you’re doing it. I feel like this starting off would be nice but [traditional] that to transition. [popups] I think that would be useful but as time goes on you need to make a judgement yourself as to if you think you are done because thats like a different skill in itself so I think that it would be useful but at the same time you need to know when to leave it alone and when you are done.	Wanted demonstration Values bright colours Values deconstructed Progression: Difficulty curve Understanding Realism:Transfer Management of cognitive load Progression: Deconstructed First Wants both Simulator performance Progression: Deconstructed First Timing of popups Learning judgement

	[Prepared?] Yes for sure [Other?] I don't know, I think its just something that I've always struggled with just that I was doing things and the machine was doing other things which is a difficult thing to teach.	Prepared for test Simulator performance
24TAU	Yes I thought it was quite good because obviously the instructions that you gave they were clear we had the opportunity to put into practice the instructions you were saying and as a group we did it all at the same time so there's no individual that has fallen behind and everyone is focussing on the same aspect of the step at the same time which was good and we also had time to focus specifically on that one aspect so we weren't focussing on say, clearing the bottom clearing the base and clearing the ADJ at the same time so that was good. [better way?] I can't think of a better way because I did quite like the way we were taught it, I don't know if that's my personal learning style but I found it very useful to be given information first and then given the opportunity to have a go in my own time then we did more of a group feedback but there's a clinician or tutor available to help if I needed it and then with the VR models there's also the opportunity to restart if you've gone wrong or see where you've gone wrong then you're able to self improve as well so I quite liked it. [decon] So I think in terms of being more realistic I think the one that we did is more realistic which is helpful in a way because when you are actually treating a patient or doing your clinical competencies next year and 3rd year you're not going to have this you will have something more resembling the tooth that we worked on so it's good because you can't baby it down too much because you have to be able to transfer those skills onto a real tooth and there can't be such a disparity with what you've practised on but I do like the idea of the gradient that you've mentioned here and here [gestures to carious spread on ex2 and enamel to be removed in ex3] because it helps the student to identify where you need to stop because obviously dentistry is, we're trying to be conservative so that approach will then really help the student see where they want to stop to not remove any healthy enamel and it seems with these first activities they are helpful for putting into perspective the instructions we were	Enjoyment Session delivery Attentional focus Session delivery Realism: Differences Concerns about transfer Hand holding Concerns about transfer Values deconstructed Values ex2 Values ex3 Understanding

	given so verbally you explained to us what the unsupported enamel is and where the ADJ is but this would help the student visualise it I think personally I was OK in understanding it verbally but I think that there was one student in our group when we were doing it who asked, who needed clarification from you where this unsupported enamel was so based on different teaching styles its obviously more useful for visual learners to do it this way but and I guess in a way audible learners can also understand it this way as well but visual learners can't really understand it just through hearing so if there was a way to just make it best for everyone as a failproof method, well not failproof, but a safer perhaps it could be with these broken down activities just to ensure that every student's understanding. I think with the gradient is good but personally I wouldn't prefer the bright colours because personally I don't need to be able to see that distinction and think it's just a little bit less realistic when you are looking at a normal tooth because if I'm so used to seeing exactly where everything is broken down, when I need to look at it on a real tooth where there's differences in morphology or different gradients that I'm not used to then I might not be able to recognise that in reality because I'm so used to seeing a broken down image. As far as which I would prefer I'm not too sure. I think there are different aspects because obviously I would prefer to work on a more realistic tooth as we did rather than little segments but then I would also prefer to see the gradients of colour that you mentioned because in theory I know there is supposed to be a gradient but on the simulators you can't see that gradient as much but you can still tell when there is no more caries in the environments. So pros and cons, pros of the way we did it, its more realistic its more sortof learn in practice which you will have to do as you go through the course anyway as in 'here you've got this tooth this is what I've told you and now put this into practice' in your own way in your own time and then you'll have a feedback session at the end so this way is much more spoon fed so personally I would prefer not having such a spoon fed approach because that's not going to be followed through the rest of the course as I'm progressing in my career so I need to be able to develop those skills myself to see what I'm given in front of me to be able to use the knowledge that I've been given to be able to put that into practise myself. As I said I do like the gradient but not the unrealistic colourings in my opinion. I think the way that we did it was a very direct approach and very thoroughly explained to us, we're	Progression: Difficulty curve Understanding Decon supports learning Understanding Dislikes bright colours Realism: Transfer Concerns about transfer Prefers TAU Prefers deconstructed Progression: Difficulty curve Hand holding Learning judgement Dislikes bright colours Attentional focus

	only focussing on a little section so it's not like remove all the caries in one go its still step by step and spoon-fed in a way but I just think that there it can't be so broken down in my opinion to such that later its say, in the 2nd year when they give us a tooth but I'm only used to looking at it like this where it's nicely separated for me then I would struggle with that transition.	Concerns about transfer
	[values realism] yes because like in 2nd year when we actually do our competencies I would get something like the models you've got over [gestures at ranking exercise models] there but the areas aren't going to be bright green or bright pink I need to be able to distinguish myself between its more subtle gradients or no gradients at all.	Realism: Differences
	<Prepared?> Yes. <Less prepared if EXP?> Hard to tell, because here you've shown me where the ADJ is [gestures at ex1] the ADJ wont be blue there but if I'd been able to understand from this and take away from it the general location of the ADJ then I would be able to put that, relate that information back onto where that isn't blue but there could be someone who for example who used to seeing these bright separated out colours and then struggle with where they don't have that direct separation so for example here if we said that part is caries and that part is healthy enamel and then they are given something like this they might not be as able to easily make the distinction between where is healthy and where the not healthy because they don't have it as separate if that makes sense.	Prepared for test Management of cognitive load Concerns about transfer
	But I think, just quickly, where you said where you're approving and disapproving different ones [ex4] I think that's quite helpful because it puts you in the perspective of what you think it should look like, sometimes you know when you're doing revision and then people say teach it to someone else because that will help you remember it, I think that's a good thing that we've done because you can see what you want to achieve by what you think is acceptable at the end so we didn't do that in the VR session that we did but it that was incorporated I think that could have helped as well.	Values ex4 Understanding
	[would like select ones?] yes to reinforce that knowledge a bit so we could do that exercise that we had but then this specific [ex4] exercise at the end so I'm saying based on what I know now what would I be expecting from someone else.	Values ex4 Understanding
	[Other?] Yeah I think vr is in general quite helpful for us because rather than focussing on passing the competency it's more able to build the skills that you	Simulator performance

	will then need to pass the competency so I think that's quite useful. I think one thing that maybe to improve is I think some of the machines give better feedback or maybe just need recalibrating I'm not sure just looking at individually because some seem to be easier to use than others but as far as using VR as a teaching tool it's been quite useful and I've enjoyed using it.	
25TAU	I would say that is a pretty effective way of teaching because you get the I think a lot that it is important to get the practical element in and it helps you understand it better by seeing it visually. If you're just getting taught that information I don't think it comes across as much in the same way, as if you're just taught these are the steps [makes a listing gesture] when you get to implement them in a way that is safe and obviously you're not going to a real patient and cause harm to them I think it is an effective in that sense. [decon] I think for this, it's probably more helpful for defining what the ADJ was because probably I would say in our session it felt focussed more on how you're supposed to clear it and how you deal with a carious lesion rather than actually understanding what it is and how far it extends and stuff like that. It was more this is just the process and this is what you are going to do rather than any kind of understanding behind it so I think this helps the understanding slightly a bit more and I think this exercise is good [ex4] because then you can identify what you think is a good one or bad one necessarily; where as in our session we were just doing what we thought was best but it may not have always necessarily been the best. I definitely feel like the other one definitely puts you in a more realistic position and I think a lot of people would get more excited by that because you actually feel like you are doing dentistry a bit more whereas this does feel more broken down and just tasks kindof. But I don't know, I kindof like the understanding behind it and use this as a starter and then move on to the more realistic things, so a bit of both I guess. I feel like it would be helpful to have the first bit actually understanding it and then moving on to doing it practically, because sometimes you can do stuff practically but not really understand what you are doing. And I feel like the understanding is quite important. I think this approach would help understand it better. [prepared?] Yes. [alt better?] I feel like our task helped	Understanding Understanding Decon supports learning Values ex4 Realism: Differences Wants both Progression: Deconstructed First Wants both Understanding Values deconstructed Learning judgement Prefers TAU

	better because although this helps with the understanding of it I guess you do need the kindof clinical exposure like where its realistic to you without it being separated for you to actually see it but I think this would help because its kindof a similar thing… I don't know. Problems? None. Other thoughts? I think that a combination would help in the most way. I feel like, because this would obviously help in terms of the understanding and the other thing would help in terms of it being real so when you came to an example you could apply both senses to it.	Realism: Transfer Prepared for test Understanding Wants both Understanding
26EXP	I think they both have their own strengths, I think I prefer the real one because sometimes like the machine is a bit laggy and it wont do what you're thinking and what you think you are putting down it will be somewhere else. I think it has a lot of potential for the future because sometimes it doesn't feel like the real thing and sometimes it does feel quite different I think you can definitely develop transferable skills doing this like hand coordination and you can work on your pen grip. [focusing on style of exercises] In that sense its quite helpful because there's a bit more colour differences compared to real life I think and its easier to orientate in VR than it is in real life and also like this section of tooth we might not get that in real life and this zoomed in. [Realism?] I think this is quite good to be honest because even if it was more realistic it wouldn't be like the real thing because its quite different so I think exercises like this is quite helpful to work on those skills even if its not like the real thing. [Prepared?] Yes I think so yes. Can't think of any changes - main issue is the lagginess [popups] yea I think it would be because everyone works at their own pace so it guides you on to the next step. [No further comments] I think it is quite good to be fair it is a bit different but its still good to work on it.	VR vs phantoms Values bright colours Understanding Realism: Transfer Simulator performance Prepared for test Simulator performance Open to popups Enjoyment

27EXP	I think the 1st session we did because it was really good because it wasn't really teeth related so it was good to learn the equipment and to just practise the motions. So that was a good start but I think this would be good to do afterwards because at that point we already had practice so we could just do fine tuned in a bit more detail and apply it to the science behind the ADJ and caries so I think that was good. Prepared? Yes I think so [TAU better?] No I think this was better because it allowed us to do it ourselves because with this you can zoom in and angle it properly and have a go and do the motions and see what happens. Its useful to see inside the tooth to start off to understand with before moving on to proper models but I think its a good starting point. [Value in whole tooth inVR] I think so because its good to get experience rather than being thrown straight in to the deep end if you go straight onto models they are just so small and you can't zoom in and really see what you are doing properly whereas this lets you see a lot more detail so I think its a good place to understand what you are doing and when you get the motions you move on to the models you have an idea what you are doing so I think its a good place to start off with that. [start with VR then move on] Reported simulator lag on ex1 but no issue [popups] I guess its personal preference but I think its always good to have a tutor explain it but maybe if the tutor explains everything at the start and then off you go then you have the messages pop up.	Progression: Difficulty curve Prepared for test Prefers deconstructed Decon supports learning Progression: Deconstructed First Progression: Difficulty curve Understanding Simulator performance Values tutor