A Review of Participant Recruitment Transparency for Sound Validation of Hip Surgery Simulators: A Novel Umbrella Approach

Authors & Affiliations

Rantilini Samaratunga, Layla Johnson, Christos Gatzidis, Ian Swain, Thomas Wainwright, Robert Middleton

Corresponding Author Details

Rantilini Samaratunga: rsamaratunga@bournemouth.ac.uk

PhD Student Department of Creative Technology, Faculty of Science and Technology, Bournemouth University*.

All Authors Details:

Layla Johnson: s5106679@bournemouth.ac.uk

Undergraduate Student, Department of Psychology, Faculty of Science and Technology, Bournemouth University*.

Prof. Christos Gatzidis: cgatzidis@bournemouth.ac.uk

Professor and Head of Creative Technology Department, Bournemouth University*

Prof. Ian Swain: iswain@bournemouth.ac.uk

Professor of Technology and Design, Bournemouth University* and of Clinical Engineering, Orthopaedic Research Institute**

Mr. Thomas W. Wainwright: twainwright@bournemouth.ac.uk

Associate Professor of Orthopaedics and Deputy Head of Orthopaedic Research Institute**

Prof. Robert Middleton: rmiddleton@bournemouth.ac.uk

Head of Orthopaedics Research Institute** and Director of Trauma at Poole Hospital***.

*Bournemouth University, Faculty of Science and Technology, Fern Barrow, Poole House, Poole, Dorset, BH12 5BB, United Kingdom

**Orthopaedic Research Institute, 6th Floor, Executive Business Centre, 89 Holdenhurst Road, Bournemouth, BH8 8EB, United Kingdom

***Poole Hospital, Longfleet Rd, Poole BH15 2JB, United Kingdom

Abstract

Malposition of implants is associated with complications, higher wear and increased revision rates in total hip replacement (THR) along with surgeon inexperience. Training THR residents to reach expert proficiency is affected by the high cost and resource limitations of traditional training techniques. Research in extended reality (XR) technologies can overcome such barriers. These offer a platform for learning, objective skill-monitoring and potentially for automated certification. Prior to their incorporation into curricula however, thorough validation must be undertaken. As validity is heavily dependent on the participants recruited, there is a need to review, scrutinise and define recruitment criteria in the absence of pre-defined standards, for sound simulator validation.

A systematic review on PubMed and IEEE databases was conducted. Training simulator validation research in fracture, arthroscopy and arthroplasty relating to the hip was included. 46 validation studies were reviewed. It was found that there was no uniformity in reporting or recruitment criteria, rendering cross-comparison challenging.

This work developed Umbrella categories to help prioritise recruitment, and formulated a detailed template of fields and guidelines for reporting criteria so that, in future, research may come to a consensus as to recruitment criteria for a hip “expert” or a “novice”.

Key Words

1. Hip
2. Simulation
3. Training
4. VR
5. Validation

Introduction and Background

Clinical Motivation

Primary and revision Total Hip Replacement (THR)[1] procedures are forecasted to increase above and beyond the aging population [1]. In 2020, the UK was expected to see a total of 143,087 THR procedures [2]. This figure is predicted to rise to 156,121 in 2030 [2]. Reducing the demand in particular for revision surgery is important with improving implant position a key objective in preventing revisions. Along with other factors, there is evidence to demonstrate that implant placement is influenced by the experience of the performing surgeon [3]. Therefore, amongst other resources, adequately trained surgeons are required to relieve the THR burden and perform procedures quickly, effectively and without complication.

Fast and effective training of surgeons however, is a complex issue. There are numerous aspects that influence the acquisition of skills and performance in surgery. Still, repetitive practice is known to be a fundamental factor. A study in Sweden showed that surgeons with higher THR case volumes had a reduction in the rates of patient accident and emergency incidents within 90 days post-surgery [4]. Chen et al. reported that when 16 high-volume surgeons were questioned, the recommended number of cases that hip trainees should be exposed to are 128 (116 ± 58.7 for hip arthroplasty specifically) during a hip fellowship [5]. The main barriers to offering surgical trainees opportunities for such high numbers of case-based practice are associated with traditional training techniques. These include the following: observation of live surgery, practice on manikins/synthetic/animal/cadaveric models and practice on real patients (increasing risk to patient safety). Such methods are often costly, heavily dependent on operating theatre (OR) availability and expert surgeon

[1] Also referred to as total hip arthroplasty (THA).

supervision, which further exacerbates resource shortages for patient procedures [6,7]. As a direct result, residents may graduate without sufficient surgical experience in vital areas [8].

Non-Traditional Training Technologies

To overcome traditional training barriers and improve patient outcomes, research has moved to using technologies such as extended reality (XR)[2] in order to design training simulators. An XR platform can offer a low-cost training environment, without compromise to patient safety. The most recent advancements in this field are towards immersive, wearable technologies, such as head-mounted displays (HMDs) coupled with controllers that allow the user more natural interaction (NI) and can render (visuo-/audio-) realistic OR environments in virtual reality (VR). This technology has shown mounting evidence of improved training in other surgical specialties, such as in general laparoscopy [9] and also in arthroscopy [10], and there is recent evidence that this XR technology can facilitate surgical training in more complex open procedures such as THR [11,12].

Simulators as Assessors

If simulators are to provide a superior alternative to traditional techniques then not only should they closely replicate the live surgical procedure in order to make training effective [13] and ensure acquisition of skills, but they should also have the objective functionality for accurate and quantitative assessment of the users' current skill-level. This would facilitate unsupervised training of the surgically naïve and, at the same time, solve the problem of subjective, variable and disparate grading systems seen cross-institutionally, cross-faculty, and cross-departmentally [14]; an issue that current accreditation bodies are particularly eager to solve [15].

[2] In this context this includes wearable head-mounted displays (HMDs) in Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR) as well as static PC-based platforms, and excludes traditional simulation such as stand-alone cadaveric, animal or synthetic bone/manikin platforms.

The digitally simulated environment offers the prime platform for implementing automated performance monitoring of skill-level for assessment and accreditation. However, whether current systems are successful in doing so requires further investigation. This issue should not be confused with the topic of much current research; whether simulators are effective training tools or not, which requires the user to show an improvement in performance [16]. The issue at hand, instead, focuses on the challenge that arises (and must be solved) prior to any attempt at evaluating training efficacy; that is whether the simulator can differentiate between users of differing skill level, most commonly an "expert" or a "novice". This is often referred to as the construct validity [17] of a simulator and its success is determined by the appropriateness and fidelity of the performance measures selected.

Vaughan et al. performed a systematic review, of mainly arthroscopic trainers, which demonstrated that simulators' in-built metrics appeared to indeed be able to discern between "expert" and "novice" technical skill levels successfully in orthopaedics [18]. More mature simulation fields, such as laparoscopy, are also taking advantage of this fact, applying machine learning techniques for intelligent assessment of trainee performance [19] and moving towards an adaptable and automated certification tool.

However, implementation of assessment technology can be quite problematic and pose a barrier in measurement selection, especially in long and complex procedures such as THR since, often, the success of the procedure is not just dependent on trainee technical psycho-motor skill proficiency [20] but also on cognitive and non-technical skills, such as decision-making [21] and leadership, which can be challenging to quantify and automatically monitor [15,20].

Furthermore, since even in psycho-motor skills acquisition there is high variability individual-to-individual, it becomes a challenging task to define, quantify and standardise criteria for classification of an "expert" and a "novice". In fact, the Dreyfus and Dreyfus

model (originally developed for pilots) [22] shows that there are in fact many skill levels in-between and categorises them as “novice”, “Advanced Beginner”, “Competent”, “Proficient” and “expert” even mentioning a modified model, which includes an ultimate level that is the “Master” level. Researchers have applied this model to the medical field and have qualitatively summarised their respective definitions [23]. Moving towards this goal in a mastery-based approach means that simulators will have to be capable of discerning more subtle skill differences, often only uncovered by current qualitative assessment techniques, for example, the Global Rating Scale [24], often used as part of the Objective Structured Assessment of Technical Skills (OSATS) in Arthroscopic Skills Assessment [25]. Therefore, it is apparent that a vigilant recruitment process must be followed, since the expertise levels of the participants will be used as the defining factor (ground truth) for the selection of quantitative measures and setting absolute proficiency standards [26], used to train these intelligent assessment simulators.

Aim and Contributions

The main aim of our work is to review how current literature has defined and reported participant recruitment criteria for hip simulator validation and also to identify if there are standards for recruitment in different specialist areas, including fracture and arthroscopy, with a particular focus however on THR. If no sufficient existing standards are found, this work aims to formulate recommendations for best practices so that future research can recruit and report participant criteria to an acceptable level for sound validation. Though there have been prior reviews conducted on orthopaedic simulators in general [27–34], to the authors’ best knowledge, this is the first review to focus solely on simulator validation for the single anatomical joint of the hip.

Materials and Methods

Systematic Search

The review process was conducted with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[3] as a template. The search query was formed by the terms listed in Table 1, and was run in two electronic databases: PubMed® and the IEEEXplore Digital Library®, and only included sources in the English language. The search was run for a 20-year period from Jan (1st) 2000 to Aug (4th) 2020.

Table 1: Search Query Terms

Joint	Procedure	Task	Platform	Subject
hip	arthroplasty	training	virtual	expert
	replacement	trainer	augmented	master
	surgery	skills	extended	surgeon
	Total	simulat*	mixed	consultant
	THA	game	reality	experienced
	THR	learn*	benchtop	specialist
	fracture	interact*	sawbones	trainees
	resurfacing	educat*	"3D-printed"	student
	preservation	teach*	three-dimensional	resident
			porcine	registrar
	arthroscopy		animal	junior
			cadaver	senior
			computer	intermediate
			PC	fellow
			manikin	novice
			phantom	beginner
				postgraduate

THA – Total Hip Arthroplasty

*** - *Wildcard Query***

[3] www.prismastatement.org/PRISMAStatement

Studies were considered eligible for inclusion within this review if they met the pre-determined inclusion and exclusion criteria listed in Table 2.

Table 2: Inclusion/Exclusion Criteria for Research in Hip Simulator Participant Recruitment

Inclusion Criteria		Exclusion Criteria
Human Participants	**Population**	Non-Human Participants
Hip Arthroplasty, Arthroscopy, Resurfacing, Fracture	**Procedure**	Any Other Anatomical Joint Soft-Tissue General Orthopaedic Skills Operations with no mention of Hip
Training Surgeons in Knowledge of Anatomy, Procedure, Psychomotor Skills, Cognitive Skills or other skills acquisition	**Task**	Pre-operative Planning Patient Rehabilitation Evaluation of a new Tool/Procedure approach Biomechanical Modelling Non-Surgical Speciality
Extended Reality, VR and AR, PC, Sawbones, 3D-Printed, Manikin, Animal, Cadaver	**Platform**	Living Human (Patient)
1) Studies that aimed to verify that the simulator was physically realistic (comparison with real surgery - face validity) 2) Studies that aimed to differentiate between "experts" and "novices" (construct validity) 3) Studies that aimed to validate efficacy of simulator for training (transfer validity)	**Study Type**	Description of System Design Only Validation not subject to human evaluation
Published in English Date of publishing 2000-2020	**Publication**	Systematic reviews Editorial commentaries

In order to be thorough, the definition of "simulation" used in this context included any research that was not an operation on a living human patient and, therefore, included any research in human cadaver training, in-vivo animal research, and synthetic bone/manikins. Even though these are classified as traditional training techniques, it was hypothesised that their recruitment processes may be able to guide those for future XR simulators.

Studies that used computer-assisted (CAS) navigation techniques with synthetic bone/manikin platforms and compared them to traditional (non-CAS synthetic bone/manikin platforms) techniques were also included, since, essentially, CAS navigation is a form of XR. Such studies are not to be confused with those that attempted to investigate whether CAS

navigation is a better technique to use in live patient procedure (these were excluded). Instead, if participants who used CAS in the training phase were then asked to perform the non-CAS procedure (and compared to other groups) for evaluation of skills acquisition, they constituted meeting our search criteria.

The search included simulators for any stage of the THR procedure and other hip-related procedures, such as arthroscopy/resurfacing/fracture-fixation/hemiarthroplasty/osteotomy, to name a few. Due to the terminology used, any general orthopaedic skills trainer with no mention of task specificity to the hip procedure or anatomy thereof (e.g. distal as opposed to proximal femur) was automatically excluded. Procedures on other joints such as the knee, shoulder, spine etc. were excluded due to a difference in the anatomical knowledge required to undertake the procedure.

Though arthroscopic procedures are included in the main review, it must be noted that the procedures utilise a very different set of skills to arthroplasty [35], and, therefore, an expert in an open procedure such as arthroplasty, might not be fairly or directly compared with that of an arthroscopic one.

The term “training” also was expected to return much research on resistance and rehabilitative training involving the patient post-surgery and, therefore, these studies were also to be excluded.

Only studies that recruited human participants were included.

The flow diagram of the search and filter process is shown in Figure 1.

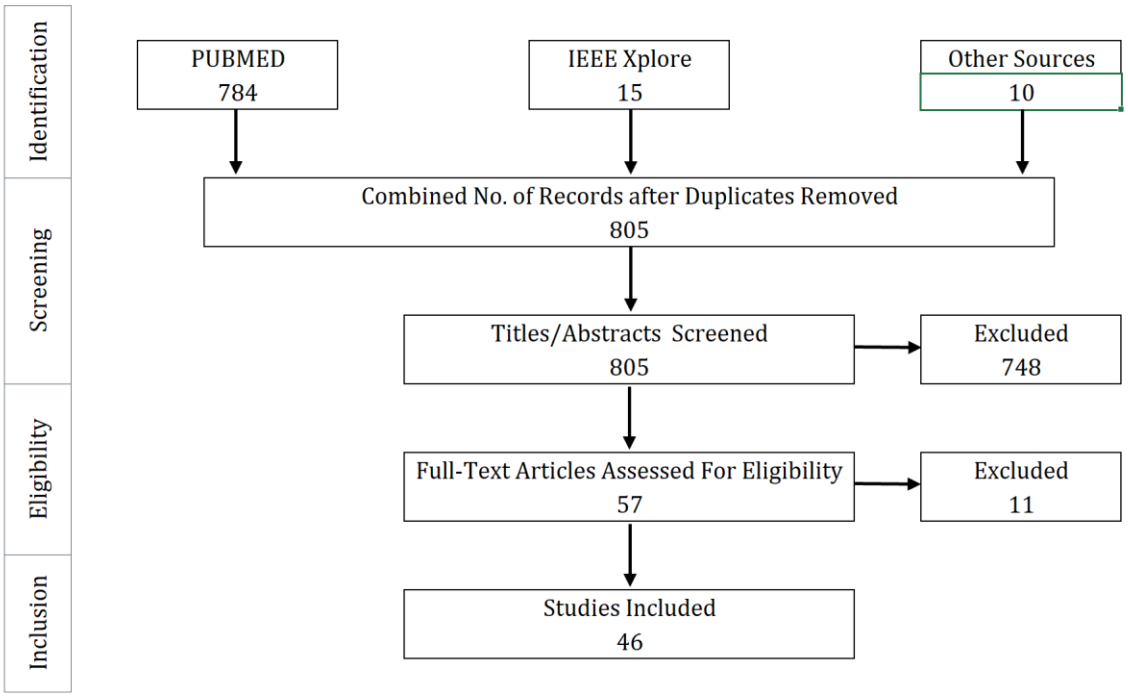

Figure 1: PRISMA Diagram

Formulation of Umbrella Categories

Numerous validation frameworks have been described in the literature but, unfortunately, the uptake of modern frameworks has been very slow in surgical simulation [36]. Borgersen et al. speculate that this could be because validation studies are not typically run by medical education specialists who may simply be using old techniques, unaware that there are updated frameworks [36]. Our work will not attempt to formulate or recommend a validation framework but, instead, attempt to create umbrella recruitment categories based on the most used terminology from publications in our literature search. It is hoped that these categories will be generic enough to prioritise recruitment criteria in such a way that the authors may formulate a detailed reporting template. The hope is that such a template may be used to make the participant recruitment process more transparent within any validation framework or research quality assessment, such as the Medical Education Research Quality Instrument (MERSQI).

Participant recruitment criteria is heavily dependent on the objective of the type of validation to be undertaken. Therefore, the authors opted to categorise the results of the systematic search into validation types that may have similar recruitment criteria. In order to do this, first, it was important to define the different types of validation. However, the literature in surgical simulation validity and respective methodology can be prone to inconsistency [36], with multiple interpretations of validation terms and the interchangeable use thereof [17,37,38].

CATEGORY_1: Face/Content/Usability

The Face validity of a simulator is often translated in the surgical field to mean a qualitative evaluation of the simulator's seemingly realistic representation of the original procedure [38], but also defined as evaluating the appropriateness of the test contents through "expert" opinion [39], which encroaches on the Content validity definition by McDougal [40]. This ambiguity makes it challenging to separate Face from Content validity[4], but the literature generally agrees; they are both very subjective forms of validation [17] that have the same goal, only differing in level of detail.

Another form of qualitative validation that participants are often recruited for, are the simulators' Usability tests. Usability is highly dependent on various factors, such as the familiarity of the participant with the interaction tools, both simulated and real, users' physical attributes that influence ergonomics, and users' individual preference and opinion, to name a few [41]. Much of these factors are contingent on those evaluated in Face and Content validation.

As a consequence of such inherent subjectivity, these types of validation have been deemed one of the most "dangerous" categories in terms of providing "seemingly" valid evidence

[4] though one literature review of orthopaedic simulators has done so by stating rigid definitions for both [28]

[42]. Therefore, if detailed and stringent recruitment criteria is set and reported, it would render the results of such tests more transparent to future researchers, who could then make independent and more informed decisions with respect to the resulting validity of such studies.

With all this in mind, the authors observed that, although it may seem advantageous to recruit participants for cohorts of differing skill levels for Category_1 validation, it is, arguably, not a necessity since "expert" opinion may be the more desirable objective. Instead, it would be more pertinent to the tests if a group of "experts" were recruited with varying physical attributes, social interests, linguistic abilities and schools of thought. Therefore, Face, Content, and Usability were grouped together to form this first umbrella category of validation, where priority is given to participants of the similar "expert" skill-level with diverse physical traits and personal backgrounds.

CATEGORY_2: Criterion/Construct

Criterion validity can have ambiguous definitions, with McDougal [40] defining it as a comparison of measurements from the simulator with the old assessment technique, and Fairhurst et al. [38] as a comparison to real or "concrete" measures. However, both agree that the subtypes of Criterion validity are termed "Concurrent validity" and "Predictive validity", with the definitions being a comparison with the current gold standard performance measurement tool and future performance predictive ability, respectively.

There is also general consensus in most surgical validation literature that Construct validity (the primary focus of this work) is the ability of the simulator to differentiate between an "expert" and a "novice" but has also been more broadly defined as differentiation of users of "differing skill levels" [40]. The latter definition is the preferred one for use in this article, as it encompasses the former. For both Criterion and Construct validity, it is compulsory that the participants to be recruited are of differing skill-level. This requirement may not seem so

obvious for Concurrent Criterion validity since one may assume that as long as participants have been (or are being) assessed by the same gold standard they can be of one skill level. However, even though the simulator may closely match the gold standard of assessment at one extremity (e.g. “novice”-level) the simulator metrics may not be as sensitive as the gold standard to subtler changes in skill level (e.g. “intermediates” and “experts”). It is obvious that for Predictive Criterion validity, differing skill levels are a must in order for automated interpolation between datasets to formulate accurate projections of skill-level, to predict performance in the OR.

Therefore, the priority requirement of Criterion and Construct validity becomes the recruitment of cohorts of differing skill levels and thus forms the second umbrella category.

CATEGORY_3: Training/Transfer

The third, and final, category aims to group all studies that look at learning curves or training effect within the simulator, as well as skills transfer[5] from the simulator to other mediums, such as cadaveric models or the OR to name a few. To assess training efficacy, the desired initial skill-level of participants would ideally be “novices”. This way, the learning curves are more detectable, whereas an “expert” is expected to have no learning curve. Training-effect and Transfer studies would typically require twin cohorts of participants with matched skill level, preferably also “novices” (so the learning curves for both cohorts can be compared), split into a control cohort and a simulation training cohort. Comparison of cohorts’ performance on a common medium after a finite training period will be fairer if these cohorts are closely matched.

[5] This category should not be confused with Predictive criterion validity, which often requires data from transfer validity studies to be able to estimate levels of skills transfer.

Therefore, the priority requirement is participants of lower skill level and matched cohort recruitment.

Result Categorisation

To summarise, the three umbrella categories used to group literature were:

- CATEGORY_1 FACE/CONTENT/USABILITY: Studies that aimed to verify that the simulator was physically realistic (comparison with real surgery, evaluate correctness/usefulness of educational content or user perception/interaction)
    - Recruitment Priority: “Experts” from diverse personal backgrounds
- CATEGORY_2 CRITERION/CONSTRUCT: Studies that aimed to differentiate between “experts” and “novices” (or differing skills levels), as well as aimed to set proficiency standards
    - Recruitment Priority: At least two cohorts of differing expertise
- CATEGORY_3 TRAINING/TRANSFER: Studies that aimed to evaluate acquisition of skills (including within simulator environment and transfer to other mediums such as cadaveric training and OR).
    - Recruitment Priority: “Novices” of similar experience

The selected sequence of the defined categories was conducted in such a way to reflect when these validity tests are usually carried out in the lifecycle of the simulator’s design and testing, i.e. only after Face/Content/Usability tests are passed should the Criterion/Construct validity take place. Subsequently, Training, and eventually, Transfer validity of the simulator should be undertaken.

Data Visualisation

Multiple software tools were used to format, sort and visualise the data gathered in this review including Jupyter Notebooks [43] along with the Plotly [44], Pandas [45,46] and NumPy [47,48] Python libraries. The data was imported from Zotero (the reference manager of choice).

Results

The results of the systematic search returned a total of 46 publications for hip simulator validity research. The majority related to Category_3 (40%), and 33.3% and 26.7% fall under Category_2 and 1, respectively.

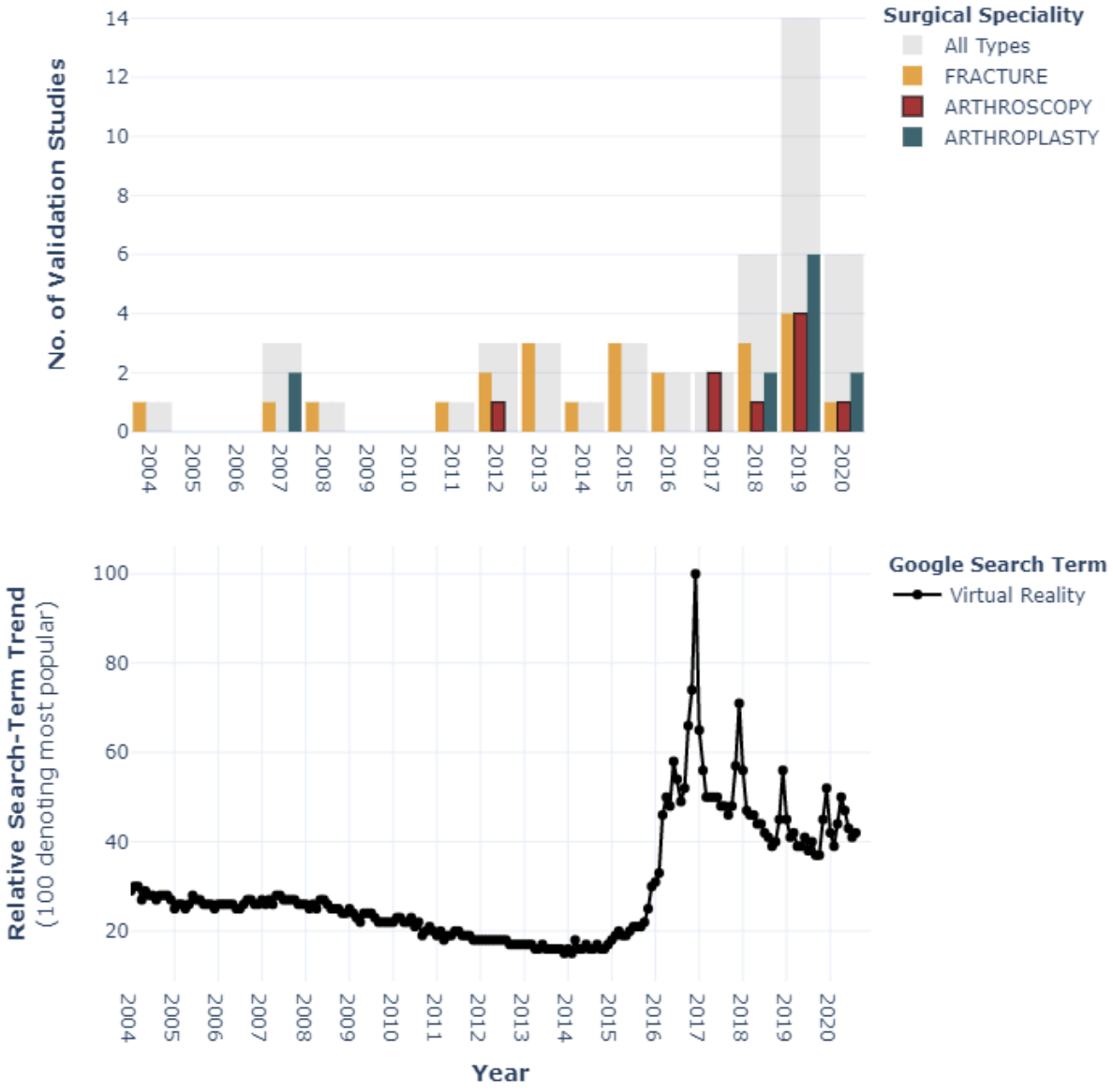

Figure 2: Top - Results of Systematic Search Sorted by Procedure Type from 2000-2020 Publications in PubMed and IEEEXplore Digital Library Databases (Plotted with Plotly[6]) [11,12,49–92]. Bottom - Worldwide Internet "VR" Search Topic Over the Past 20 Years with Peak Interest in December 2016 (Data from Google Trends[7])

[6] Plotly Technologies Inc. Collaborative data science. Montréal, QC, 2015. https://plot.ly
[7] https://trends.google.com/trends/explore?date=all&q=%2Fm%2F07_ny

The general trend in hip surgery simulators can be seen in Figure 2, with Fracture simulators dominating the 2000s and 2010s and arthroscopy and arthroplasty only gaining popularity in the past few years.

The earliest instance of Category_3 validation in arthroplasty using a surgical simulator was seen in 2007 by Gofton et al. [52] and Cobb et al [50]. Many previous reviews did not include these studies, most likely due to the use of the terminology CAS navigation, which can be misleading. However, adding computer assistance to a traditional simulator platform, such as synthetic bones, and evaluating its training and transfer effect to another platform [50,52] did indeed qualify it as a simulator for inclusion in our study, as it falls under the XR domain.

The first hip arthroscopy simulator validation seen was in 2012 by Pollard et al. [55] and this too was for a bench-top manikin. It was not until five years later that any further arthroscopy research was conducted. Due to their use of physical models, these studies likely suffered from the disadvantages of traditional training methods mentioned in section 0. This could explain why no more research was seen in this area until modern XR technologies became a more viable/cost-efficient option by around late 2016, demonstrated by the sudden spike in popularity of the Google search-term for "virtual reality" (bottom plot in Figure 2). It was also during this time that Vaughan et al.'s review of orthopaedic VR simulators highlighted the lack of development in hip and THR simulators [32].

We observed (see Figure 3) that there was a dearth of research in hip arthroplasty under Category_2, with only one study by Logishetty et al. in January 2020 [88] that examined the performance of "experts" vs. "residents" in THR even though Construct validity was not explicitly stated as the aim of the study. There were also only two studies published with respect to Category_1 validation in arthroscopy [76,80].

7 studies having their sole objective of Category_1 validation did not recruit participants in multiple cohorts and therefore were grouped together in

Table 3 [51,69,75,77–79,84].

18 studies were grouped based on their similar recruitment, with multiple cohorts of differing expertise for Category_2 (see

Author	Tillander et al.	Blythe et al.	Froelich et al.	Riehl & Widmaier	Pedersen et al.	Akhtar et al.	Kho et al.	Dwyer et al.	Khanduja et al.	Phillips et al.	Christian et al.	Erturan et al.	Sugand et al.	Gustafsson et al.	Bauer et al.	Cychosz et al.	Gallagher et al.	Logishetty et al.
Year	2004	2008	2011	2012	2014	2015	2015	2016	2017	2017	2018	2018	2018	2019	2019	2019	2019	2020
Cohort Name	"Medical Students"	"MSs"	"Group 1"	"Novice Operator"	"Group 1"	"Novice"	"Novice"	"PGY-1"	"Novice"	"Novice"	"Medical Students"	"Novices"	"Novices"	"Novices"	"Non-Expert"	"Novice"	"Novice"	"Ortho. Residents"
Qualitative Description	Medical Students	Medical Students	Junior Residents PGY-1 & PGY-2 (in month 11 of academic year). All PGY-1s had been exposed to/assisted in DHS or similar	X-Ray Technicians	Untrained Novices - Ortho. Interns Without Prior Hip Fracture Exp.	PGY-1 to PGY-4	Minimal to No Surg. Exp. PGY-1; UG Eng.	PGY-1 Residents at end of 1st year of orthopaedic training	Surgical residents at various stages of 6-year speciality orthopaedic training less than 250 cases	Junior Resident PGY1-3; Senior Resident PGY4-5	Medical Students with Ortho. Interest	Medical Students without prior hip arthroscopy experience; Interns with minimal operative exposure	Postgraduate Orthopaedic Trainees Without experience of sim before PGY-2 to PGY-5 Less than 10 Hip Cases	In 1st year of specialisation	Orthopaedic residents and board-certified orthopaedic surgeons without prior experience in hip arthroscopy	Less than 25 previous cases. Ortho. Surgery Residents PGY-1; Medical students interested in orthopaedic surgery	Performing or assisting in less than 50 procedures	PGY-1 to PGY-4
Participants (n)	15	6	6	6	10	10	6; 24	9	10	27; 10	15	10; 10	8	38	33	4; 4	16	32
Surgical Training/Experience (yrs)	**0	0	**Range: 0.92-1.92	**0		**Range: 0-4	**Range: 0-0.5	**0-1	**Range: 0-6	**0.5 - 2.5; **3.5 - 4.5	Mean (SD): 0 (0)	**0; Unknown	**Range: 1 - 5	Median (Range): **0.583 (0-1.833) working in orthop. Dept.		**0-1; 0		**Range: 0-4
Procedure-Specific Full Cases Performed (n)				?	0	Mean: 1			<10		Mean (SD): 0 (0)	**Median (Range): 0 (0-0); 0(0-0)	< 10	Median (Range): 1 (0-10)	0	Mean +- SD (Range): 0 +- 0 (0)	<50	0
Procedure-Specific Partial Cases Performed (n)						Mean: 2				Mean: 0 0.78 0.15; 1.86 8.33							<50	0
Procedure-Specific Cases Assisted (n)										Mean: 0.4 1.78 2.61; 8.29 22				?				0
Procedure-Specific Cases Observed (n)				?		Mean: 6												
Any Surgical Simulator Use											0%			yes/no: 17/21				
Gaming Experience		Fort-nightly									66.67%							Never Before, Once, Occasionally, Monthly, At Least Weekly: 20,5,1,5,1

Sim Platform-Specific Use (E.g. PC, VR etc.)	Median (Range) %: 66 (40-90)								0
Cohort Name	"BTs"	"Group 2"	"Inter-mediate"	"PGY-4"	"Resi-dents"	"Trainee"	"Inter-mediates"	"Inter-mediate"	
Qualitative Description	Basic Trainees; *some fixing as many hip fractures as advanced trainees*	Senior Residents PGY 3 to PGY-5 in 11th month	PGY-5 to PGY-12	PGY-4 Residents who have undertaken 6month trauma rotation and arthroplasty rotations	Ortho. Residents PGY-2 to PGY-5	Orthopaedic residents of varying seniority; Knee/Shoulder fellowship trainees	Post Graduate Orthopaedic PGY4 to PGY9 of 10 to 30 cases	25-74 cases. Junior residents (PGY 2/3); Senior resident (PGY4/5)	
Participants (n)	6	9	10	9	17	18; 10	7	8; 1	
Surgical Training/Experience (yrs)	<3	**Range: 2-5	**Range: 4-12		Mean (SD): 3.5 (1.2)		**Range: 3-9	**1-3; 3-5	
Procedure-Specific Full Cases Performed (n)			Mean: 66		Mean (SD): 2 (35)	Median (Range): 0 (0-0); 0 (0-0)	Range: 10-30	Mean +- SD (Range): 4 +- 5.68 (0-15)	
Procedure-Specific Partial Cases Performed (n)			Mean: 28						
Procedure-Specific Cases Assisted (n)									
Procedure-Specific Cases Observed (n)			Mean: 26						
Any Surgical Simulator Use					88.24%				
Gaming Experience	Monthly				70.59%				

Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 37 (14-80)														
Cohort Name	"Exper-ienced"	"ATs"	"Ortho. Surgeons"	"Group 2"	"Expert"	"Exper-ienced"	"Expert"	"Exper-ienced"	"Fellows /Attend-ings"	"Consult-ant"	"Experts"	"Experts"	"Expert"	"Exper-ienced"	"Expert"	"Expert"
Qualitative Description	Experienced Ortho. Surgeons	Advanced Trainees	Residents & Attendings	Ortho. Surgeons - senior residents or specialists	>PGY-12	Ortho. Residents PGY-3, PGY-4; Faculty surgeons who take general call	Faculty members - practicing orthopaedic surgeons completed 6-year training and certified as orthopaedic specialist more than 250 cases, all but 1 performed more than 500	Sports Medicine Fellows; Staff Surgeons	Attending Ortho. Surgeons & Ortho. Trauma fellows	Expert hip arthroscopists	PGY7 or more with more than 40 cases	Consultant surgeons specialized in orthopaedic trauma surgery	All but 1 have completed more than 200 arthroscopic procedures. All completed more than 150	Senior residents (PGY4/5) ; fellowship trained orthopaedic sports medicine attending (prof) more than 74 procedures		"Expert hip surgeons"
Participants (n)	10	6	6, 4	10	10	5; 5	9	5; 5	18	4	11	8	9	3; 1	6	4
Surgical Training/Experience (yrs)		>4			**>11	**>2	Mean +- SD: 20.1 +- 4.7		Mean (SD): 10.6 (6.2)		**>6	Median (Range): 4 (3-15) as full-time orthopaedic traumatologist		**3-5; **5++		
Procedure-Specific Full Cases Performed (n)			?	>20	427		>250		212	Median (Range): 527 (209-800)	>40		>150	Mean +- SD (Range): 52 +- 85.74 (1-200)	>=50	
Procedure-Specific Partial Cases Performed (n)					234			Mean: 6.4 ; 99							>=50	
Procedure-Specific Cases Assisted (n)								Mean: 16.0; 166								
Procedure-Specific Cases Observed (n)			?		229											
Any Surgical Simulator Use									50%							
Gaming Experience		Monthly							66.67%							
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 41(20-66)														

) [49,53,54,56,61–63,66–68,70,71,73,76,80,81,86,88]

In total there were 24 studies that investigated the training effect and/or transfer of skills to some extent, with the majority (62.5%) being in Fracture, 29.2% in Arthroplasty and only two (8.3%) in Arthroscopy. 21 of these studies had the primary goal of determining simulator training usefulness based on Category_3 validation and hence reported recruitment similarly [11,12,50,52,55,57–60,64,65,72,74,82,83,85,87,89–92].

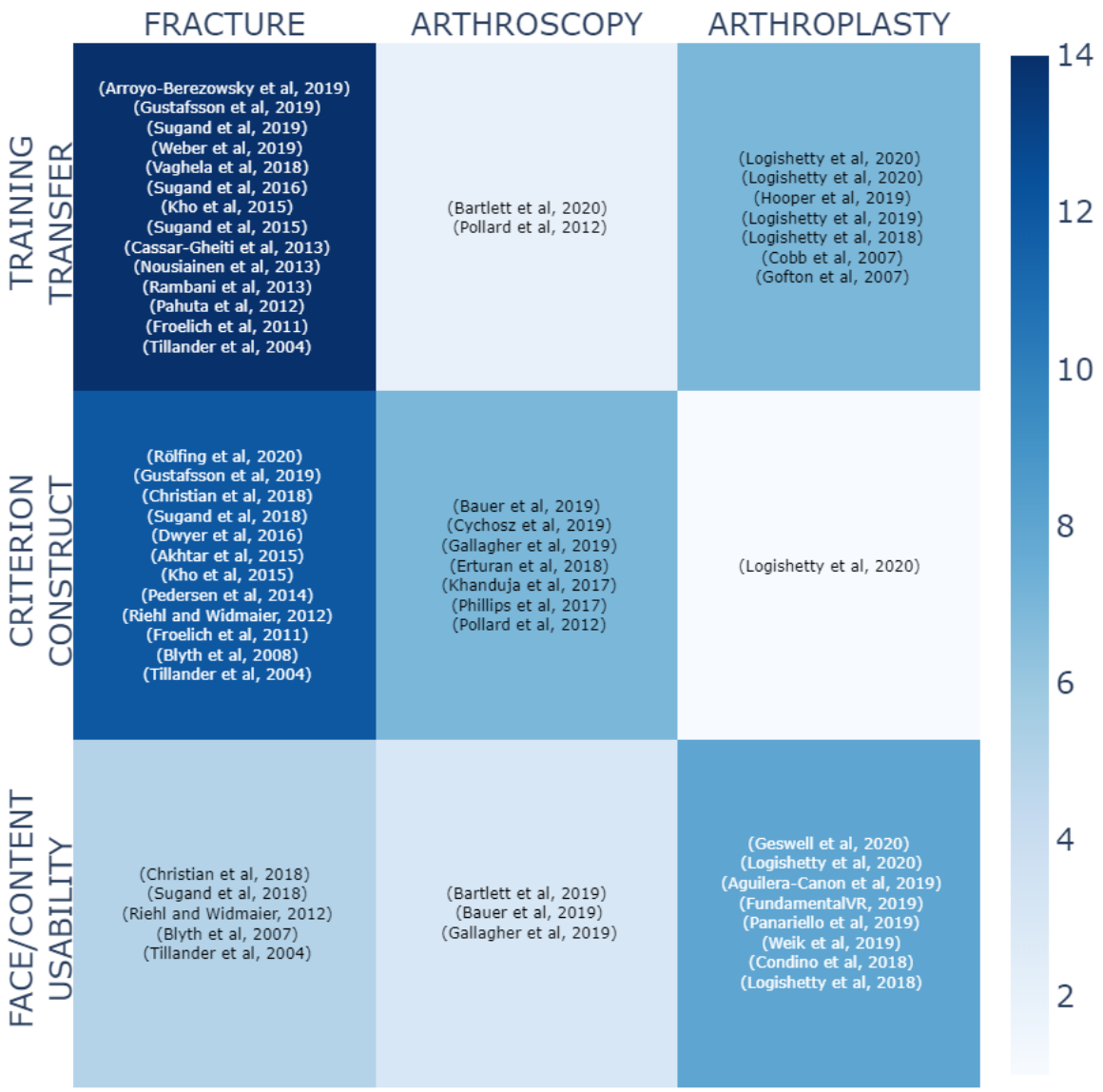

Figure 3: Heatmap of Results of Systematic Search Categorised by Procedure Type (x-axis) and Novel Umbrella recruitment Category (y-axis). Colour scale correlates to number of publications in that category, with darker shades denoting higher numbers [11,12,49–92], plotted with Plotly

Table 3: Reported Participant Recruitment Criteria for Studies with single objective of Category_1 Validation

Author	Blythe et al.	Condino et al.	Bartlett et al.	Panariello e al.	FundamentalVR	Aguilera-Canon et al.	Weik et al.
Year	2007	2018	2019	2019	2019	2019	2019
Surgical Speciality	FRACTURE	ARTHROPLASTY	ARTHROSCOPY	ARTHROPLASTY	ARTHROPLASTY	ARTHROPLASTY	ARTHROPLASTY
Qualitative Description	3 x Medical Students, 4 x junior trainees and 3x fellows/orthopaedic surgeons	Engineers; med staff: students, orthop. Residents, and orthop. Surgeons	7 x Faculty Members (Orthopaedic Surgeons) programme. 18x orthopaedic surgical residents from a variety of training programmes across Europe	2 Volunteers of healthy and robust constitution	1 x Medical Student, 3 x FY2, 1 x CT1, 7 x CT2, 1 x ST1, 4 x ST2, 6 x ST3, 4 x ST4, 6 x ST5, 2 x ST6, 3 x ST7, 2 x ST8	Hip surgeons and surgical trainees	Orthopaedic Surgeons
Participants (n)	10	10;10 (6,1,3)	25	2	40	7	2
Age (yrs)	20-50	23:48:32:7 (Min Max Mean STD)		23, 40			
Sex/Gender (Male: Female)	9:1	13:7:0 (Male, Female, Non-Binary)		2:0			2:0
Handedness (Right:Left)		18:2:0 (right, left, ambidextrous)		2:0			
Vision (10/10 naked eyes, lens-corrected to 10/10)		10,10		No deficiency in visual or haptic perception			
Colour Blindness (no, yes)		20, 0					
English Reading		0,0,12,8 (none, limited, familiar, experienced)					
English Speaking		0,2,11,6 (none, limited, familiar, experienced)					
Procedure-Specific Full Cases Performed (n)						prior knowledge of task	
Sim Platform-Specific Use (E.g. PC, VR etc.)	0 in VR, PC Knowledge median = 3.9/10	experience with AR: 8,5,5,2 (None, limited, familiar, experienced) experience with MS HoloLens: 16,3,1,0 (None, limited, familiar, experienced)	No Participants had previous experience using virtual reality of any kind				

Table 4: Reported Participant Recruitment Criteria for Studies with at least one objective as Category_2 Validation – Condensed Table

Author	Tillander et al.	Blythe et al.	Froelich et al.	Riehl & Widmaier	Pedersen et al.	Akhtar et al.	Kho et al.	Dwyer et al.	Khanduja et al.	Phillips et al.	Christian et al.	Erturan et al.	Sugand et al.	Gustafsson et al.	Bauer et al.	Cychosz et al.	Gallagher et al.	Logishetty et al.
Year	2004	2008	2011	2012	2014	2015	2015	2016	2017	2017	2018	2018	2018	2019	2019	2019	2019	2020
Cohort Name	"Medical Students"	"MSs"	"Group 1"	"Novice Operator"	"Group 1"	"Novice"	"Novice"	"PGY-1"	"Novice"	"Novice"	"Medical Students"	"Novices"	"Novices"	"Novices"	"Non-Expert"	"Novice"	"Novice"	"Ortho. Residents"
Qualitative Description	Medical Students	Medical Students	Junior Residents PGY-1 & PGY-2 (in month 11 of academic year). All PGY-1s had been exposed to/assisted in DHS or similar	X-Ray Technicians	Untrained Novices - Ortho. Interns Without Prior Hip Fracture Exp.	PGY-1 to PGY-4	Minimal to No Surg. Exp. PGY-1; UG Eng.	PGY-1 Residents at end of 1st year of orthopaedic training	Surgical residents at various stages of 6-year speciality orthopaedic training less than 250 cases	Junior Resident PGY1-3; Senior Resident PGY4-5	Medical Students with Ortho. Interest	Medical Students without prior hip arthroscopy experience; Interns with minimal operative exposure	Postgraduate Orthopaedic Trainees Without experience of sim before PGY-2 to PGY-5 Less than 10 Hip Cases	In 1st year of specialisation	Orthopaedic residents and board-certified orthopaedic surgeons without prior experience in hip arthroscopy	Less than 25 previous cases. Ortho. Surgery Residents PGY-1; Medical students interested in orthopaedic surgery	Performing or assisting in less than 50 procedures	PGY-1 to PGY-4
Participants (n)	15	6	6	6	10	10	6; 24	9	10	27; 10	15	10; 10	8	38	33	4; 4	16	32
Surgical Training/Experience (yrs)	**0	0	**Range: 0.92-1.92	**0		**Range: 0-4	**Range: 0-0.5	**0-1	**Range: 0-6	**0.5 - 2.5; **3.5 - 4.5	Mean (SD): 0 (0)	**0; Unknown	**Range: 1 - 5	Median (Range): **0.583 (0-1.833) working in orthop. Dept.		**0-1; 0		**Range: 0-4
Procedure-Specific Full Cases Performed (n)				?	0	Mean: 1			<10		Mean (SD): 0 (0)	**Median (Range): 0 (0-0); 0(0-0)	< 10	Median (Range): 1 (0-10)	0	Mean +- SD (Range): 0 +- 0 (0)	<50	0
Procedure-Specific Partial Cases Performed (n)						Mean: 2				Mean: 0 0.78 0.15; 1.86 8.33							<50	0
Procedure-Specific Cases Assisted (n)										Mean: 0.4 1.78 2.61; 8.29 22				?				0
Procedure-Specific Cases Observed (n)				?		Mean: 6												
Any Surgical Simulator Use											0%			yes/no: 17/21				
Gaming Experience		Fort-nightly									66.67%							Never Before, Once, Occasionally, Monthly, At Least Weekly: 20,5,1,5,1
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range) %: 66 (40-90)																0

Cohort Name	"BTs"	"Group 2"	"Inter-mediate"	"PGY-4"	"Resi-dents"	"Trainee"	"Inter-mediates"	"Inter-mediate"
Qualitative Description	Basic Trainees; *some fixing as many hip fractures as advanced trainees*	Senior Residents PGY 3 to PGY-5 in 11th month	PGY-5 to PGY-12	PGY-4 Residents who have undertaken 6month trauma rotation and arthroplasty rotations	Ortho. Residents PGY-2 to PGY-5	Orthopaedic residents of varying seniority; Knee/Shoulder fellowship trainees	Post Graduate Orthopaedic PGY4 to PGY9 of 10 to 30 cases	25-74 cases. Junior residents (PGY 2/3); Senior resident (PGY4/5)
Participants (n)	6	9	10	9	17	18; 10	7	8; 1
Surgical Training/Experience (yrs)	<3	**Range: 2-5	**Range: 4-12		Mean (SD): 3.5 (1.2)		**Range: 3-9	**1-3; 3-5
Procedure-Specific Full Cases Performed (n)			Mean: 66		Mean (SD): 2 (35)	Median (Range): 0 (0-0); 0 (0-0)	Range: 10-30	Mean +- SD (Range): 4 +- 5.68 (0-15)
Procedure-Specific Partial Cases Performed (n)			Mean: 28					
Procedure-Specific Cases Assisted (n)								
Procedure-Specific Cases Observed (n)			Mean: 26					
Any Surgical Simulator Use					88.24%			
Gaming Experience	Monthly				70.59%			
Sim Platform-Specific Use (E.g. PC, VR etc.)	Median (Range): 37 (14-80)							

Cohort Name	"Exper-ienced"	"ATs"	"Ortho. Surgeons"	"Group 2"	"Expert"	"Exper-ienced"	"Expert"	"Exper-ienced"	"Fellows /Attend-ings"	"Consult-ant"	"Experts"	"Experts"	"Expert"	"Exper-ienced"	"Expert"	"Expert"
Qualitative Description	Experienced Ortho. Surgeons	Advanced Trainees	Residents & Attendings	Ortho. Surgeons - senior residents or specialists	>PGY-12	Ortho. Residents PGY-3, PGY-4; Faculty surgeons who take general call	Faculty members - practicing orthopaedic surgeons completed 6-year training and certified as orthopaedic specialist more than 250 cases, all but 1 performed more than 500	Sports Medicine Fellows; Staff Surgeons	Attending Ortho. Surgeons & Ortho. Trauma fellows	Expert hip arthroscopists	PGY7 or more with more than 40 cases	Consultant surgeons specialized in orthopaedic trauma surgery	All but 1 have completed more than 200 arthroscopic procedures. All completed more than 150	Senior residents (PGY4/5) ; fellowship trained orthopaedic sports medicine attending (prof) more than 74 procedures		"Expert hip surgeons"
Participants (n)	10	6	6, 4	10	10	5; 5	9	5; 5	18	4	11	8	9	3; 1	6	4
Surgical Training/Experience (yrs)		>4			**>11	**>2	Mean +- SD: 20.1 +- 4.7		Mean (SD): 10.6 (6.2)		**>6	Median (Range): 4 (3-15) as full-time orthopaedic traumatologist		**3-5; **5++		
Procedure-Specific Full Cases Performed (n)			?	>20	427		>250		212	Median (Range): 527 (209-800)	>40		>150	Mean +- SD (Range): 52 +- 85.74 (1-200)	>=50	
Procedure-Specific Partial Cases Performed (n)					234			Mean: 6.4 ; 99							>=50	
Procedure-Specific Cases Assisted (n)								Mean: 16.0; 166								
Procedure-Specific Cases Observed (n)			?		229											
Any Surgical Simulator Use									50%							
Gaming Experience		Monthly							66.67%							
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 41(20-66)														

Key

** Inferred from qualitative description

* Figures for all cohorts collectively (no breakdowns given)

Discussion

In 2010, Van Nortwick et al. [37] provided a general guideline for validation methodology for surgical simulators, including a generic table of recommendations for subject demographic reporting. However, the results of our literature search showed that only one subsequent hip simulator validity study stated following this guideline, and that the majority do not state following any validation framework at all. This was shown to be a general problem in all areas of surgical simulation in a review of validation methods conducted in 2018 by Borgersen et al. [36]. It outlined that studies were becoming more vague at reporting what validation frameworks they followed, if any, and even using validation nomenclature incorrectly, with a very notable increase of such publications from only 2.1% in 2012 to 23.3% in 2017 [36].

To avoid this for the relatively newer field of THR surgical simulators, we have reviewed the current validation research in prior hip surgery simulation and proceeded to analyse them under novel umbrella categories.

CATEGORY_1 Face/Content/Usability

Until very recently, XR simulators for training of open hip procedures, especially in THR, did not exist [32], which accounts for why validation research in this area is sparse. This is mainly due to the nature of the surgical procedure. THR or arthroplasty in general (unlike arthroscopy), has demanding technical visualisation requirements, including depth-rendering requirements, as the surgeon is no longer visualising patient anatomy through a 2D camera or radiograph (as in arthroscopy or fluoroscopy). Therefore, more sophisticated visual display techniques and positional controls are required for a larger reachable workspace and a more accurate representation of the user's avatar in the virtual environment. As these factors are

dependent on the user's unique physical attributes, e.g. arm-span, it can be challenging to design user-agnostic interaction.

Technology has now advanced sufficiently to be able to provide realistic visual patient rendering, as FundamentalVR[8] demonstrate in their hardware agnostic platform, recently accredited by the Royal College of Surgeons [77]. More and more simulators are starting to emerge from commercial companies such as OssoVR[9] ORamaVR[10] and Virtamed[11], though thorough evaluation for all these platforms is currently sparse. Figure 2 shows this general trend in hip surgery simulators over the past 20 years, the most advanced simulator research being in hip fracture-fixation, and the slow shift to hip arthroplasty as XR platform technology has advanced (its popularity highlighted by Figure 2's Google search-term trend plot), with an approximate two-to-three year lag before these technology trends are integrated into simulators and then observed in validation research.

However, as much as visual realism has advanced, haptic realism appears to still be lacking. Participants have often fed back that there is much still required in simulator haptics even in MIS procedures [49,73,76,80], let alone in high-force arthroplasty procedures. Only two studies in hip fracture had participants reach some consensus that force realism had been achieved [56,82] and one study in arthroplasty that the instruments used felt realistic [74]. These are likely due to the simulators in these trials being physical-model based rather than virtual-environment based.

Almost all studies in this umbrella category had the outcome that the simulator was deemed a "useful" tool for surgical education, usually when the interaction with the platform was seen as intuitive [69,75]. Studies like [77] solely used participants who were still in training

[8] https://www.fundamentalvr.com/
[9] https://ossovr.com/
[10] https://oramavr.com/
[11] https://www.virtamed.com/en/

(ranging from Medical Students all the way to final year surgical specialty years), raising the question as to how researchers should distinguish between opinions to disregard and ones to pursue, in order to make the constructive design changes necessary to improve the simulator. Of course, if the participant is regarded a field "expert" by his peers, then it is possible his opinion would carry more weight [81] than a "novice" who may be classified as not having performed any operations (i.e. Medical Students) or is still in training.

An early Face validity study by Blyth et al. in 2007 for a PC-based hip Fracture simulator showed that the majority (56%) of participants in fact disagreed with the need for haptics to be incorporated [51]. However, upon closer inspection, the demographics of the participants consisted of three Medical Students, four Junior Trainees and only three Fellows/Orthopaedic surgeons; participants in the first two categories most-likely[12] never having performed the procedure on a live patient. Since there was no reporting as to whom exactly disagreed, it is difficult to soundly use such validation results in any design/implementation decisions; regardless of the questionnaire or rating method used. Since then, it has been reported that incorporation of haptic feedback improves general surgical VR simulator training effect [93].

Condino et al.'s participant demographic reporting was, by far, the most extensive amongst all the studies that only conducted validation in Category_1 [69] (see

[12] Amount of experience is unknown as the study does not report the number of prior cases participants have performed/assisted/observed

Table 3). Affinity for this style of numerical data tabularisation might stem from the researchers' predominantly engineering-based backgrounds, as opposed to clinical-based ones. It was also the only study to include ambidexterity and non-binary as options for handedness and gender respectively [69]. However, the use of engineers to evaluate the Usability of the AR THR simulator that is primarily meant for surgeons should be questioned, and was, it is likely, carried out in order to increase group size (and hence power analysis), which can be common practice amongst surgical simulator validation studies. According to consensus guidelines by Carter et al., even "expert" opinions are only classed as a low Level of Evidence (LoE) of 4, compared to Randomized controlled trials of "good" quality and of adequate sample size, constituting a LoE of 1b [94]. However, as discussed, the majority of hip simulator validation studies, whose main goal is Category_1 validation, do not reach this low LoE since mixed experience-level and, in some cases, no "expert" participants were used [95] (see

Table 3).

Furthermore, only a few studies in this category mentioned following multi-institutional [49,77,82,84,87] recruitment of participants, meaning that the majority of studies cannot be taken as institutionally agnostic since, even if "experts" were used, their opinions may differ from one institution to another [14].

The self-published FundamentalVR study was able to recruit participants from the most number of training levels (from Medical Student all the way to ST8 - Specialty Training Year 8[13]) since their recruitment took place at a conference [77]. This is a testament to the benefits of more portable and cloud-based simulators, since, in the future, physical and geographical locations will no longer be an issue for participant recruitment as it is hoped technological, as well as engagement-level, barriers for take-home training kits will be reduced [96]. However, in this instance, recruitment for the Face validation of the posterior THR simulator was dependent on the attendees of a conference and therefore numbers of participants for each skill-level were very low.

Therefore, it is apparent that for publications in this category, though "expert" opinion is vital, "expert" recruitment is lacking, and understandably so; availability of consultant surgeons is often limited. However, if a small number of diverse and leading "experts" are recruited, this could, arguably, better support the utility of conducting Category_1 validation than the opinions of higher numbers of participants of mixed skill-level.

To ensure researchers do not overlook the importance of diversity, the authors have added the Physical/Personal Attributes fields to the recommended guideline (Table 5) to aid in the reporting and recruitment for Category_1 validation studies. Attributes such as eye colour and ethnicity have never before been reported in any previous studies, but should, in our view, be included since eye-tracking and hand-tracking technologies may differ in accuracy due to these wearer attributes (143,144), hence affecting the user experience and therefore opinions towards the simulator. Since these technologies are increasingly being

[13] https://www.rcseng.ac.uk/careers-in-surgery/trainees/foundation-and-core-trainees/surgery-career-paths/

incorporated in the latest commercial/mainstream HMDs (HTC VIVE Pro Eye[14], Microsoft HoloLens 2[15] etc.), until the technologies have been perfected, this data should be gathered in any Category_1 validation and can be used by engineers and technologists to improve such devices, if need be.

The matter of ensuring that the participants are of "expert" skill-level however is much more challenging and the authors aim to provide further guidelines to aid Category_2 validation.

CATEGORY_2 Criterion/Construct

The thorough justification behind undertaking Construct validation for a simulator is often overlooked. In fact, Cook [99] contends that the simulator's ability to discriminate between expert and novice cohorts adds little to the validity of the simulator and should not be the only form of validation that occurs. However, the inability to differentiate may immediately highlight issues before other stages of validation [99]. Our work therefore assumes that Construct validity is only one part of the simulation validation process and should be carried out along with Criterion and Transfer studies [99] before it can be used as evidence of usefulness. Furthermore, stringent participant recruitment processes should address all the confounding issues stated in [99].

However, as is the case in this work, if a primary goal of the simulator is to serve as an assessment (pass/fail) and certification tool, then Category_2 validation must be carried out to such a sound level, that any proficiency standards [100] set within such simulators should be beyond reproach, since, eventually[16], the simulator could serve as the final frontier before trainees are then allowed to perform patient procedures.

[14] https://www.vive.com/uk/product/vive-pro-eye/overview/
[15] https://www.microsoft.com/en-us/hololens
[16] depending on the simulator's place in the training curricula

Factoring out the Learning curve of the simulator

For true assessment, the simulator used should not in any way influence the participant's current skill level, whether it is through extensive time, and hence practice, on the platform beforehand, or providing feedback on performance metrics during the experiment. [54] and [73] made a point to mention that no feedback was visible to participants. However, it is not clear as to how much time participants should be allowed in order to familiarise themselves with the simulator beforehand. Earlier studies [49,53] had "experts" guide participants through the first operation. Some papers have allowed no practice time at all with only visual [62,70], audio-visual [68], written instruction [67], or standardized explanation/demonstration [56,73,82] of each task. Other studies allowed two minutes or less [54,80], five minutes or less [76] and even up to ten minutes [63], the latter recognising that this long familiarisation time could indeed have affected their findings.

As it is known that repeated practice does facilitate skills acquisition [101], ideally, participants should only be given a written, audio or video description of finite time and have no practice on the simulator itself. If the interaction perfectly mirrored that of a live surgical procedure over all five sensory domains, then, zero prior practice would not be an issue. However, though this may be better achieved in the arthroscopic procedures, this is not the case for arthroplasty. Researchers must be able to decouple the simulator's learning curve from that of the user's baseline skill level.

The expert cohort is the ideal place to start. Since we can assume their knowledge and skill level of the actual procedure is complete, any learning curve observed based on the outcomes of the simulated procedure can be assumed to be their familiarisation with the simulator. Gustafsson et. al found that "experts" took as long as "novices" for familiarisation with a hip fracture osteosynthesis simulator and surmised it was due to the simulator not allowing them to perform the task as they would in real surgery, and therefore had to change their approach to operate within the simulation environment [81].

It may even take longer for the "experts" to familiarise themselves with the tools, due to differences in age and hence personal exposure to digital platforms than "novices", and researchers have tried to adjust for this through questionnaires about prior personal gaming experience (even though research has not yet

found strong correlation between skills acquisition and gaming [102]). Unfortunately, there was no consensus on the type of scale used for recording this attribute, with some studies using amount of time (1 hour per week etc.) and others qualitative Likert scales (e.g. frequently, less-frequently etc.).

However, if the longer time is used as the upper limit for number of attempts/time, we can safely use this baseline of number of attempts (or time taken) as the cut-off for the familiarisation phase for all the "novice" participants during the study. Whatever the familiarisation task however, it must not be related to the surgical procedure, but should be more a game-based task, so as not to improve the user's surgical knowledge or psycho-motor skill level.

Therefore, the better approach would be to invite "experts" to take part before the "novice" cohort, but how can one define an "expert" or "novice"?

Expert or Novice?

As the metrics chosen when setting proficiency standards are heavily dependent on how skilled the participants are, this gives rise to the question of defining quantitative criteria for an "expert", "intermediate" or a "novice", so they may be recruited accordingly. Sadideen et al. reviewed whether surgeons' innate talent is what leads them to become "experts" in their field, but reported that though it is a factor, such expertise is attained through significant practice [103]. Surgeons are required to keep a logbook during training and throughout their career, in which the number of cases performed, assisted-with and observed are marked down. This case history can serve as a quantitative measure for the amount of prior practice and therefore skill-level. Though Chen et al.'s subjective survey attempted to quantify the number of cases required for adequate proficiency [5], it is unclear how many prior cases constitute an "expert" in hip surgery fields.

To ascertain this criteria, we have compiled

Author	Tillander et al.	Blythe et al.	Froelich et al.	Riehl & Widmaier	Pedersen et al.	Akhtar et al.	Kho et al.	Dwyer et al.	Khanduja et al.	Phillips et al.	Christian et al.	Erturan et al.	Sugand et al.	Gustafsson et al.	Bauer et al.	Cychosz et al.	Gallagher et al.	Logishetty et al.
Year	2004	2008	2011	2012	2014	2015	2015	2016	2017	2017	2018	2018	2018	2019	2019	2019	2019	2020
Cohort Name	"Medical Students"	"MSs"	"Group 1"	"Novice Operator"	"Group 1"	"Novice"	"Novice"	"PGY-1"	"Novice"	"Novice"	"Medical Students"	"Novices"	"Novices"	"Novices"	"Non-Expert"	"Novice"	"Novice"	"Ortho. Residents"

Qualitative Description	Medical Students	Medical Students	Junior Residents PGY-1 & PGY-2 (in month 11 of academic year). All PGY-1s had been exposed to/assisted in DHS or similar	X-Ray Technicians	Untrained Novices - Ortho. Interns Without Prior Hip Fracture Exp.	PGY-1 to PGY-4	Minimal to No Surg. Exp. PGY-1; UG Eng.	PGY-1 Residents at end of 1st year of orthopaedic training	Surgical residents at various stages of 6-year speciality orthopaedic training less than 250 cases	Junior Resident PGY1-3; Senior Resident PGY4-5	Medical Students with Ortho. Interest	Medical Students without prior hip arthroscopy experience; Interns with minimal operative exposure	Postgraduate Orthopaedic Trainees Without experience of sim before PGY-2 to PGY-5 Less than 10 Hip Cases	In 1st year of specialisation	Orthopaedic residents and board-certified orthopaedic surgeons without prior experience in hip arthroscopy	Less than 25 previous cases. Ortho. Surgery Residents PGY-1; Medical students interested in orthopaedic surgery	Performing or assisting in less than 50 procedures	PGY-1 to PGY-4
Participants (n)	15	6	6	6	10	10	6; 24	9	10	27; 10	15	10; 10	8	38	33	4; 4	16	32
Surgical Training/Experience (yrs)	**0	0	**Range: 0.92-1.92	**0		**Range: 0-4	**Range: 0-0.5	**0-1	**Range: 0-6	**0.5 - 2.5; **3.5 - 4.5	Mean (SD): 0 (0)	**0; Unknown	**Range: 1 - 5	Median (Range): **0.583 (0-1.833) working in orthop. Dept.		**0-1; 0		**Range: 0-4
Procedure-Specific Full Cases Performed (n)				?	0	Mean: 1			<10		Mean (SD): 0 (0)	**Median (Range): 0 (0-0); 0(0-0)	< 10	Median (Range): 1 (0-10)	0	Mean +- SD (Range): 0 +- 0 (0)	<50	0
Procedure-Specific Partial Cases Performed (n)						Mean: 2				Mean: 0 0.78 0.15; 1.86 8.33							<50	0
Procedure-Specific Cases Assisted (n)										Mean: 0.4 1.78 2.61; 8.29 22				?				0
Procedure-Specific Cases Observed (n)				?		Mean: 6												
Any Surgical Simulator Use											0%			yes/no: 17/21				
Gaming Experience		Fort-nightly									66.67%							Never Before, Once, Occasionally, Monthly, At Least Weekly: 20,5,1,5,1
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range) %: 66 (40-90)																0
Cohort Name		"BTs"	"Group 2"			"Inter-mediate"		"PGY-4"			"Resi-dents"	"Trainee"	"Inter-mediates"			"Inter-mediate"		

Qualitative Description		Basic Trainees; *some fixing as many hip fractures as advanced trainees*	Senior Residents PGY 3 to PGY-5 in 11th month		PGY-5 to PGY-12		PGY-4 Residents who have undertaken 6month trauma rotation and arthroplasty rotations			Ortho. Residents PGY-2 to PGY-5	Orthopaedic residents of varying seniority; Knee/Shoulder fellowship trainees	Post Graduate Orthopaedic PGY4 to PGY9 of 10 to 30 cases				25-74 cases. Junior residents (PGY 2/3); Senior resident (PGY4/5)		
Participants (n)		6	9		10		9			17	18; 10	7				8; 1		
Surgical Training/Experience (yrs)		<3	**Range: 2-5		**Range: 4-12					Mean (SD): 3.5 (1.2)		**Range: 3-9				**1-3; 3-5		
Procedure-Specific Full Cases Performed (n)					Mean: 66					Mean (SD): 2 (35)	Median (Range): 0 (0-0); 0 (0-0)	Range: 10-30				Mean +- SD (Range): 4 +- 5.68 (0-15)		
Procedure-Specific Partial Cases Performed (n)					Mean: 28													
Procedure-Specific Cases Assisted (n)																		
Procedure-Specific Cases Observed (n)					Mean: 26													
Any Surgical Simulator Use										88.24%								
Gaming Experience		Monthly								70.59%								
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 37 (14-80)																
Cohort Name	"Exper-ienced"	"ATs"		"Ortho. Surgeons"	"Group 2"	"Expert"	"Exper-ienced"		"Expert"	"Exper-ienced"	"Fellows /Attend-ings"	"Consult-ant"	"Experts"	"Experts"	"Expert"	"Exper-ienced"	"Expert"	"Expert"

Qualitative Description	Experienced Ortho. Surgeons	Advanced Trainees	Residents & Attendings	Ortho. Surgeons - senior residents or specialists	>PGY-12	Ortho. Residents PGY-3, PGY-4; Faculty surgeons who take general call	Faculty members - practicing orthopaedic surgeons completed 6-year training and certified as orthopaedic specialist more than 250 cases, all but 1 performed more than 500	Sports Medicine Fellows; Staff Surgeons	Attending Ortho. Surgeons & Ortho. Trauma fellows	Expert hip arthroscopists	PGY7 or more with more than 40 cases	Consultant surgeons specialized in orthopaedic trauma surgery	All but 1 have completed more than 200 arthroscopic procedures. All completed more than 150	Senior residents (PGY4/5) ; fellowship trained orthopaedic sports medicine attending (prof) more than 74 procedures		"Expert hip surgeons"
Participants (n)	10	6	6, 4	10	10	5; 5	9	5; 5	18	4	11	8	9	3; 1	6	4
Surgical Training/Experience (yrs)		>4			**>11	**>2	Mean +- SD: 20.1 +- 4.7		Mean (SD): 10.6 (6.2)		**>6	Median (Range): 4 (3-15) as full-time orthopaedic traumatologist		**3-5; **5++		
Procedure-Specific Full Cases Performed (n)			?	>20	427		>250		212	Median (Range): 527 (209-800)	>40		>150	Mean +- SD (Range): 52 +- 85.74 (1-200)	>=50	
Procedure-Specific Partial Cases Performed (n)					234			Mean: 6.4 ; 99							>=50	
Procedure-Specific Cases Assisted (n)								Mean: 16.0; 166								
Procedure-Specific Cases Observed (n)			?		229											
Any Surgical Simulator Use									50%							
Gaming Experience		Monthly							66.67%							
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 41(20-66)														

, demonstrating how some publications in Category_2 have reported recruitment demographics and what criteria they have set.

For the validation category with the most import placed on knowing the ground-truth and absolute skill-level, it was interesting to see that there is no uniformity amongst reporting participant recruitment details, especially for the characteristic that is perhaps most pertinent to construct validity; prior case history. Approximately a third of studies (6 out of the 18) do not detail prior case history and recorded years of experience [49,53,54,56,63,66]. Unfortunately, case history cannot be inferred from these years of training since there are cross-institutional differences in curricula, meaning trainees are exposed to different amounts of training mediums and live patient procedures.

It was anticipated that the more recent studies published may be more thorough at reporting participant demographic details. It was also expected that post-2010 work would attempt to adhere to the framework described by Van Nortwick et al. and Schout et al. [17,37]. Intriguingly, this was not the case in the latter instance and only one study by Bauer et al. [80] in this category quoted following any framework at all; the Van Nortwick framework. However, there was a marginal increase in reporting prior procedures, though no uniformity in the number of cases that would constitute an “expert” or “novice”. In addition to Borgerson et al.’s reasoning for this trend in study methodologies, we speculate this is because of the following; though generic validation framework papers and publications on simulator design [104] and validation [105] processes are available, they do not provide specific tabular templates for participant recruitment reporting with prioritisation of which fields are most important for Category_2 validation.

Phillips et al. were one of the most thorough at reporting the case history of their participants recruited for their arthroscopic hip label repair synthetic bone simulator [68]. Prior case histories of hip, knee and shoulder arthroscopy procedures were recorded and higher cases were found to correlate to better performance on the simulator [68]. This indicates that there are similar cross-anatomy psycho-motor skills in arthroscopy, and, therefore, prior case history of other anatomical joint arthroscopies could be used as a criterion to determine the participant’s skill level in hip arthroscopy. This is in keeping with previous evidence that demonstrate laparoscopic skills are transferable to arthroscopy and vice versa [62].

However, it is not clear whether knee and shoulder arthroplasty skills are transferable to hip arthroplasty, as none of the arthroplasty studies in Category_2 have attempted to report or recruit based on participants' prior case histories for other anatomical joints. It is also not evident whether prior hip arthroscopic case numbers correlate to performance in arthroplasty of the same anatomical joint.

The authors included the "prior experience" fields in Table 5 by combining all those seen in Category_2 literature (

Author	Tillander et al.	Blythe et al.	Froelich et al.	Riehl & Widmaier	Pedersen et al.	Akhtar et al.	Kho et al.	Dwyer et al.	Khanduja et al.	Phillips et al.	Christian et al.	Erturan et al.	Sugand et al.	Gustafsson et al.	Bauer et al.	Cychosz et al.	Gallagher et al.	Logishetty et al.
Year	2004	2008	2011	2012	2014	2015	2015	2016	2017	2017	2018	2018	2018	2019	2019	2019	2019	2020
Cohort Name	"Medical Students"	"MSs"	"Group 1"	"Novice Operator"	"Group 1"	"Novice"	"Novice"	"PGY-1"	"Novice"	"Novice"	"Medical Students"	"Novices"	"Novices"	"Novices"	"Non-Expert"	"Novice"	"Novice"	"Ortho. Residents"
Qualitative Description	Medical Students	Medical Students	Junior Residents PGY-1 & PGY-2 (in month 11 of academic year). All PGY-1s had been exposed to/assisted in DHS or similar	X-Ray Technicians	Untrained Novices - Ortho. Interns Without Prior Hip Fracture Exp.	PGY-1 to PGY-4	Minimal to No Surg. Exp. PGY-1; UG Eng.	PGY-1 Residents at end of 1st year of orthopaedic training	Surgical residents at various stages of 6-year speciality orthopaedic training less than 250 cases	Junior Resident PGY1-3; Senior Resident PGY4-5	Medical Students with Ortho. Interest	Medical Students without prior hip arthroscopy experience; Interns with minimal operative exposure	Postgraduate Orthopaedic Trainees Without experience of sim before PGY-2 to PGY-5 Less than 10 Hip Cases	In 1st year of specialisation	Orthopaedic residents and board-certified orthopaedic surgeons without prior experience in hip arthroscopy	Less than 25 previous cases. Ortho. Surgery Residents PGY-1; Medical students interested in orthopaedic surgery	Performing or assisting in less than 50 procedures	PGY-1 to PGY-4
Participants (n)	15	6	6	6	10	10	6; 24	9	10	27; 10	15	10; 10	8	38	33	4; 4	16	32
Surgical Training/Experience (yrs)	**0	0	**Range: 0.92-1.92	**0		**Range: 0-4	**Range: 0-0.5	**0-1	**Range: 0-6	**0.5 - 2.5; **3.5 - 4.5	Mean (SD): 0 (0)	**0; Unknown	**Range: 1 - 5	Median (Range): **0.583 (0-1.833) working in orthop. Dept.		**0-1; 0		**Range: 0-4
Procedure-Specific Full Cases Performed (n)				?	0	Mean: 1			<10		Mean (SD): 0 (0)	**Median (Range): 0 (0-0); 0(0-0)	< 10	Median (Range): 1 (0-10)	0	Mean +- SD (Range): 0 +- 0 (0)	<50	0
Procedure-Specific Partial Cases Performed (n)						Mean: 2				Mean: 0 0.78 0.15; 1.86 8.33							<50	0
Procedure-Specific Cases Assisted (n)										Mean: 0.4 1.78 2.61; 8.29 22				?				0
Procedure-Specific Cases Observed (n)				?		Mean: 6												
Any Surgical Simulator Use											0%			yes/no: 17/21				

Gaming Experience	Fort-nightly				66.67%			Never Before, Once, Occasionally, Monthly, At Least Weekly: 20,5,1,5,1
Sim Platform-Specific Use (E.g. PC, VR etc.)	Median (Range) %: 66 (40-90)							0
Cohort Name	"BTs"	"Group 2"	"Inter-mediate"	"PGY-4"	"Resi-dents"	"Trainee"	"Inter-mediates"	"Inter-mediate"
Qualitative Description	Basic Trainees; *some fixing as many hip fractures as advanced trainees*	Senior Residents PGY 3 to PGY-5 in 11th month	PGY-5 to PGY-12	PGY-4 Residents who have undertaken 6month trauma rotation and arthroplasty rotations	Ortho. Residents PGY-2 to PGY-5	Orthopaedic residents of varying seniority; Knee/Shoulder fellowship trainees	Post Graduate Orthopaedic PGY4 to PGY9 of 10 to 30 cases	25-74 cases. Junior residents (PGY 2/3); Senior resident (PGY4/5)
Participants (n)	6	9	10	9	17	18; 10	7	8; 1
Surgical Training/Experience (yrs)	<3	**Range: 2-5	**Range: 4-12		Mean (SD): 3.5 (1.2)		**Range: 3-9	**1-3; 3-5
Procedure-Specific Full Cases Performed (n)			Mean: 66		Mean (SD): 2 (35)	Median (Range): 0 (0-0); 0 (0-0)	Range: 10-30	Mean +- SD (Range): 4 +- 5.68 (0-15)
Procedure-Specific Partial Cases Performed (n)			Mean: 28					
Procedure-Specific Cases Assisted (n)								
Procedure-Specific Cases Observed (n)			Mean: 26					
Any Surgical Simulator Use					88.24%			

Gaming Experience		Monthly							70.59%							
Sim Platform-Specific Use (E.g. PC, VR etc.)		Median (Range): 37 (14-80)														
Cohort Name	"Exper-ienced"	"ATs"	"Ortho. Surgeons"	"Group 2"	"Expert"	"Exper-ienced"	"Expert"	"Exper-ienced"	"Fellows /Attend-ings"	"Consult-ant"	"Experts"	"Experts"	"Expert"	"Exper-ienced"	"Expert"	"Expert"
Qualitative Description	Experienced Ortho. Surgeons	Advanced Trainees	Residents & Attendings	Ortho. Surgeons - senior residents or specialists	>PGY-12	Ortho. Residents PGY-3, PGY-4; Faculty surgeons who take general call	Faculty members - practicing orthopaedic surgeons completed 6-year training and certified as orthopaedic specialist more than 250 cases, all but 1 performed more than 500	Sports Medicine Fellows; Staff Surgeons	Attending Ortho. Surgeons & Ortho. Trauma fellows	Expert hip arthroscopists	PGY7 or more with more than 40 cases	Consultant surgeons specialized in orthopaedic trauma surgery	All but 1 have completed more than 200 arthroscopic procedures. All completed more than 150	Senior residents (PGY4/5) ; fellowship trained orthopaedic sports medicine attending (prof) more than 74 procedures		"Expert hip surgeons"
Participants (n)	10	6	6, 4	10	10	5; 5	9	5; 5	18	4	11	8	9	3; 1	6	4
Surgical Training/Experience (yrs)		>4			**>11	**>2	Mean +- SD: 20.1 +- 4.7		Mean (SD): 10.6 (6.2)		**>6	Median (Range): 4 (3-15) as full-time orthopaedic traumatologist		**3-5; **5++		
Procedure-Specific Full Cases Performed (n)			?	>20	427		>250		212	Median (Range): 527 (209-800)	>40		>150	Mean +- SD (Range): 52 +- 85.74 (1-200)	>=50	
Procedure-Specific Partial Cases Performed (n)					234			Mean: 6.4 ; 99							>=50	
Procedure-Specific Cases Assisted (n)								Mean: 16.0; 166								
Procedure-Specific Cases Observed (n)			?		229											
Any Surgical Simulator Use									50%							
Gaming Experience		Monthly							66.67%							

Sim Platform-Specific Use (E.g. PC, VR etc.)	Median (Range): 41(20-66)

), to further research in hip arthroplasty simulator validation. The only way to uncover relationships of cross-anatomy and cross-procedure (arthroplasty and arthroscopy) psycho-motor skills transfer is through the thorough reporting of prior case numbers for hip arthroscopy and knee/shoulder arthroplasty.

CATEGORY_3 Training/Transfer

In order to ascertain whether a trainee is improving (and hence plotting the learning curve), the objective metrics selection should be sound. When designing proficiency standards, measurable metrics must be chosen which can be monitored both in the training platform as well as the transferred platform. This is one of the major barriers for THR, as unlike arthroscopy, open surgery proves to be more challenging for tool path monitoring etc. Hence, the final umbrella category is heavily dependent on the sound completion of Category_2 validation.

Assuming Category_1 and _2 validation have been soundly conducted, the only recruitment barriers that are left at this stage pertain to the close matching of cohorts. This problem however is not unique to training simulator validation and has been looked at extensively in the medical field, as it parallels patient recruitment in clinical trials for control cohorts. Category_3 recruitment often does not suffer from lack of cohort sizes either, because in order to ensure learning curves are detectable, the desired participants should be at the novice-level. Participants at this skill level are much more readily available than those at the opposite extreme. However, in order to ensure close matching of the cohorts, transparency of participant demographic is vital. Therefore, the authors hope that the fields in Table 5, formulated from publication analyses in Category_1 and _2, can serve as a tool to simplify this matching process.

Pahuta et al. conducted the earliest study in objectively measuring cognitive skills acquisition for learning acetabular fracture anatomy. This does not fully report participant demographics; only training year and sex were recorded but not reported [57]. The work also mentions conducting Vandenberg et al.'s visuospatial ability assessment [106], pre-empting some correlation with performance, but these scores were also not reported, and were also found not to have any statistical significance.

Arroyo-Berezowsky et al. do provide a detailed breakdown of all participants including institution, PG year, gender and age. They, like Pahuta et al., did not find any statistical significance between gender and performance [83]. This tends to be the common trend amongst more recent work by Logishetty et al. as well [88].

However, by including visuo-spatial ability and the sex/gender field in the reporting guidelines (Table 5) it serves as a reminder for future researchers as to the importance of recording inherent participant ability to a much higher degree, even though this is more useful for Category_1 studies.

Recommendations

In order to recruit appropriate participants to correctly perform all the categories of validity testing, it must first be understood who the simulator is catering to. It is important whether its use is for the earlier training, more advanced, late-stage training or as just-in-time [104,107] training to refresh already qualified surgeons who are returning to work (i.e. have not performed the procedure after a while), or indeed a combination of all. Therefore, during the training needs analysis (TNA) described by Schout et al. [17], it is important to also perform stakeholder analysis [108] to ascertain starting requirements for participant recruitment.

In the majority of validation studies found in this review, there were omissions in reporting the surgical simulator's intended place in the training curriculum and hence declaring what

surgical training years/level of experience would ideally be considered for novice-level and who would ideally be considered an expert.

The authors recommend that, once the simulator's place is established in the training scheme (i.e. before, after or in conjunction with other training platforms such as dry-bones) and the stakeholders are identified, the example table laid out in Table 5 should be used to recruit and report participant demographic data. In instances where field data cannot be gathered, researchers should explicitly state why this was not possible (e.g. to maintain anonymity of participants etc.) and should still include these fields but simply leave them blank or 0:0 (e.g. in eye dominance), so that there is a uniformity when it comes to analysis and comparison of reports.

Few studies have looked at the transference of arthroscopy to arthroplasty, or knee arthroplasty to hip arthroplasty. Therefore, completing these fields (of related or similar procedures) will allow future studies to find correlations or relationships for cross-anatomy skills transfer; a gap in the current understanding of surgical psycho-motor skills acquisition.

Surgical training and research in COVID-19

The global lockdown as a result of the COVID-19 pandemic has forced many surgical training and educational institutions to re-think their methods for delivery and access to knowledge and learning [109,110]. As surgical residents are being called upon to fill the shortage of non-surgical healthcare professionals [111], it might seem that they have reduced opportunities and time for training. However, one study showed that residents who performed soft-tissue surgical procedures had more opportunities to perform emergency operations and use open approaches otherwise reserved for special cases [112]. Still, as elective surgeries are put on hold, this means that most orthopaedic residents will not be exposed to live arthroscopic and arthroplasty procedures. Therefore, to avoid skills decay and to further unsupervised learning, the importance of the rapid adoption of XR technologies in training

programmes is apparent [113,114]. The template produced in this work hopes to standardise recruitment in such a way in order to accelerate validity research so simulators may be incorporated into training curriculums in the very near future.

Table 5: Recommended standardisation for reporting participant criteria in cohorts for all Umbrella categories

No.	Field	Further Description	Example Recommended Reporting
01	**Cohort Name**	Qualitative Name	For Category_2: More Experienced, Less Experienced
02	**Cohort Description**	Qualitative description of the cohort, preferably with industry standard code names. The order stated here can be taken	Medical Students, PGY1, PGY2, PGY3 through to CT1, ST2 etc.
03	**Total No. of Participants in Cohort**	This should be total no. after sufficient elimination [100]	(n)
04	**No. of centres/ countries recruited from**	Not recommended to display actual centre names to preserve anonymity. Stating country instead, can give a clue regarding curriculum diversity	(n) Institutes A,B and C Country using alpha-2 country code e.g. GB for the U.K., LK for Sri Lanka etc.
05	**No. of Participants recruited from each centre**	To see if there are correlations between centres and performance	State as a ratio (n:n:n)

06	**No. of Participants in Training Year Code Name**	State as a ratio in the order stated in field 02	PGY1, PGY2, PGY3, CT1, ST2 etc. n:n:n:n:n:n:n
07	**No. of Years in Surgical Training/Practice**	Should include months as well, since someone at the start of the academic year may not have the same experiences as someone at the end of the year. This should only include Surgical Training, so Medical students should be listed as having 0 yrs of surgical training	1.5 years (1 year and 6 months) for a PGY2 student. For a consultant surgeon, it should include the no. of years in training and the no. of years practicing after training completed combined

Prior Experience:

The following should be recorded for only the past 5 years, to take into account frequency and hence skill decay

08	**No. of procedure-specific cases performed**	This is only referring to live cases on a living patient	Mean, Median, Mode, Range +- SD
09	**No. of procedure-specific cases assisted in or partially performed**	This is only referring to live cases on a living patient	Mean, Median, Mode, Range +- SD
10	**No. of similar cases performed**	This is only referring to live cases on a living patient	Mean, Median, Mode, Range +- SD

11	**No. of similar cases assisted in or partially performed**	This is only referring to live cases on a living patient	Mean, Median, Mode, Range +- SD
12	**No. of procedure-specific cases observed**	This is only referring to live cases on a living patient	Mean, Median, Mode, Range +- SD
13	**No. of procedure-specific simulated cases performed**	As simulator research advances, what research work may not foresee is that the same surgeons might have already been part of a study with a simulator platform	Mean, Median, Mode, Range +- SD
14	**No. of similar simulated cases performed**	As simulator research advances, what researcher work may not foresee is that the same surgeons might have already been part of a study with a simulator platform	Mean, Median, Mode, Range +- SD
15	**Amount of time spent on any simulator (minutes)**	Difficult to say no. of times used, so amount of time is instead asked, and can be compared to no. of procedure-specific cases (because approx. length of procedure is known).	Mean, Median, Mode, Range +- SD

		Qualitative reporting methods avoided, to make retrospective data entry and analysis easier.	
		Personal/Physical Attributes:	
16	**Age (yrs)**	This can be used to find correlations between technology/platform proficiency	Mean, Median, Mode, Range +- SD
17	**Sex at birth and/or Gender**	The term gender has been used by too many studies, when in fact it is sex (pertaining to the biological and physical attributes of the individual) that may be more pertinent for Category_1 recruitment. Both Sex and Gender may play a more significant role in learning and hence Category_3 recruitment [115], therefore both have been included in this field. Condino et al. attempted to be further inclusive by providing the option for “Non-Binary”	State as a ratio (Male:Female:Non-Binary)
18	**Height (cm)**	If not known prior to experiment, should be measured. Useful for user experience when	Mean, Median, Mode, Range +- SD

		setting up VR headset height/table-top height etc.	
19	**Handedness**	Best reported as a ratio of no. of participants who have their dominant hand as their right or left (or both)	State as a ratio (Right:Left:Ambidextrous)
20	**Ethnicity**	For estimating skin pigmentation and whether negative interaction experiences correlates with bad hand tracking due to differing infra-red reflectance	Standard List can be taken from any other database
21	**Hand/Glove Size (cm)**	Mean of both hands Useful for hand-tracking interaction	Mean, Median, Mode, Range +- SD
22	**Eye Dominance**	Especially useful for eye-tracking simulator platforms	State as a ratio (Right:Left)
23	**Colour Blindness**	Could affect perception of color-coded visual cues (e.g., red warning regions and green safe zones) in simulator	State as a ratio (Yes:No)
24	**Require Corrective Lenses (External)**	Especially useful for eye-tracking simulator platforms, and can influence opinion of comfort in wearable headsets	State as a ratio (Yes:No)

25	**Eye Colour**	Especially useful for eye-tracking simulator platforms, and can influence opinion if tracking performs differently for differing eye colours	State as a ratio (Colour1:Colour2:Colour3) Can be stated as Red Green Blue (RGB) values from 0 to 255 depending on if tracking technology has colour measurement functionality
26	**Visuo-spatial Ability**	To pre-determine innate individual ability. Important to report for Category_3 Umbrella for purposes of matching cohorts	As scores from the three dimensional rotations test [106]

Limitations

We recognize that this study was only conducted on two databases and only for hip-related simulators. Therefore, to understand relationships for related surgeries, there needs to be more research conducted on other anatomical joints as well. This work does not do a deep review of all validation frameworks, therefore cannot provide detailed recommendations for experiment design, but only guidelines for reporting and selecting participant recruitment criteria.

Conclusion

The recruitment of participants for Construct validity of simulators has the potential for multiple relationships to be uncovered. As surgeons (in-training or otherwise) are few and far between, and their availability to take part in such studies can invariably be quite limited, maximum data collection must therefore be carried out in order to facilitate research in surgical simulators. With thorough and transparent reporting of participant recruitment, more sound validation can be achieved for hip surgery simulators.

This work attempted to identify best practices during the participant recruitment process. A novel approach for prioritising recruitment objectives through umbrella categorisation of validation types was undertaken. It was found that hip research employed multiple cohorts of apparent differing skill-levels without often stating exact recruitment criteria for each. For the research that has, often the recruitment criteria is based on years of experience rather than number of cases; this needs to be scrutinised if simulators are to become true assessors of surgical skill in hip procedures. Literature in this area also suffers from non-uniformity of reporting recruitment criteria. A standardised table of reporting participant recruitment criteria was devised to facilitate cross-comparison of studies, and to prevent researchers from relaxing recruitment criteria for the purposes of larger cohort sizes. Instead, the authors would envisage that the use of the standardised template can facilitate cross-institutional

collaboration for larger and more diverse cohorts, whilst simultaneously maintaining stringent and transparent recruitment criteria.

Acknowledgements

This research was undertaken as part of project THRiLS, which is part-funded by Dorset Creative. Many thanks to Louise Burgess and Tikki Immins from the Orthopaedic Research Institute at BU for helping with the PRISMA process.

References

[1] NHFD 2019 annual report [Internet]. [cited 2020 Jan 12]. Available from: https://www.nhfd.co.uk/20/hipfractureR.nsf/docs/2019Report.

[2] Pabinger C, Lothaller H, Portner N, et al. Projections of hip arthroplasty in OECD countries up to 2050. HIP International. 2018;28:498–506.

[3] Meek RMD, Allan DB, McPhillips G, et al. Epidemiology of Dislocation after Total Hip Arthroplasty. Clinical Orthopaedics and Related Research®. 2006;447:9–18.

[4] Jolbäck P, Rolfson O, Cnudde P, et al. High annual surgeon volume reduces the risk of adverse events following primary total hip arthroplasty: a registry-based study of 12,100 cases in Western Sweden. Acta Orthopaedica. 2019;90:153–158.

[5] Chen AW, Steffes MJ, Laseter JR, et al. The education and training of future hip preservation surgeons: aggregate recommendations of high-volume surgeons. J Hip Preserv Surg. 2018;5:307–311.

[6] Burke E, Felle P, Crowley C, et al. Augmented reality EVAR training in mixed reality educational space. 2017 IEEE Global Engineering Education Conference (EDUCON). 2017. p. 1571–1579.

[7] Boylan MR, Perfetti DC, Naziri Q, et al. Is Orthopedic Department Teaching Status Associated With Adverse Outcomes Of Primary Total Hip Arthroplasty? The Journal of Arthroplasty. 2017;32:S124–S127.

[8] Stirling ER, Lewis TL, Ferran NA. Surgical skills simulation in trauma and orthopaedic training. J Orthop Surg Res [Internet]. 2014 [cited 2020 Feb 5];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4299292/.

[9] Larsen CR, Oestergaard J, Ottesen BS, et al. The efficacy of virtual reality simulation training in laparoscopy: a systematic review of randomized trials. Acta Obstetricia et Gynecologica Scandinavica. 2012;91:1015–1028.

[10] Aïm F, Lonjon G, Hannouche D, et al. Effectiveness of Virtual Reality Training in Orthopaedic Surgery. Arthroscopy. 2016;32:224–232.

[11] Logishetty K, Rudran B, Cobb JP. Virtual reality training improves trainee performance in total hip arthroplasty: a randomized controlled trial. The Bone & Joint Journal. 2019;101-B:1585–1592.

[12] Hooper J, Tsiridis E, Feng JE, et al. Virtual Reality Simulation Facilitates Resident Training in Total Hip Arthroplasty: A Randomized Controlled Trial. The Journal of Arthroplasty. 2019;34:2278–2283.

[13] Brown D. The Role of Simulation in the Learning of Surgical Skills. Bulletin. 2013;95:292–295.

[14] Sobel AD, Hartnett D, Hernandez D, et al. Global variability in orthopedic surgery training. Orthop Rev (Pavia) [Internet]. 2019 [cited 2020 Feb 5];11. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6776924/.

[15] Kowalewski TM, Lendvay TS. Performance Assessment. In: Stefanidis D, Korndorffer Jr. JR, Sweet R, editors. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties [Internet]. Cham: Springer International Publishing; 2019 [cited 2019 Jun 13]. p. 89–105. Available from: https://doi.org/10.1007/978-3-319-98276-2_9.

[16] Johnston MJ, Paige JT, Aggarwal R, et al. An overview of research priorities in surgical simulation: what the literature shows has been achieved during the 21st century and what remains. The American Journal of Surgery. 2016;211:214–225.

[17] Schout BMA, Hendrikx AJM, Scheele F, et al. Validation and implementation of surgical simulators: a critical review of present, past, and future. Surg Endosc. 2010;24:536–546.

[18] Vaughan N, Dubey VN, Wainwright TW, et al. Can virtual-reality simulators assess experience and skill level of orthopaedic surgeons? 2015 Science and Information Conference (SAI). 2015. p. 105–108.

[19] Winkler-Schwartz A, Bissonnette V, Mirchi N, et al. Artificial Intelligence in Medical Education: Best Practices Using Machine Learning to Assess Surgical Expertise in Virtual Reality Simulation. Journal of Surgical Education. 2019;76:1681–1690.

[20] Atesok K, Hurwitz S, Anderson DD, et al. Advancing Simulation-Based Orthopaedic Surgical Skills Training: An Analysis of the Challenges to Implementation [Internet]. Advances in Orthopedics. 2019 [cited 2020 Jan 31]. Available from: https://www.hindawi.com/journals/aorth/2019/2586034/.

[21] Clever SL, Novack DH, Cohen DG, et al. Evaluating surgeons' informed decision making skills: pilot test using a videoconferenced standardised patient. Med Educ. 2003;37:1094–1099.

[22] Batalden P, Leach D, Swing S, et al. General competencies and accreditation in graduate medical education. Health Aff (Millwood). 2002;21:103–111.

[23] Patnaik R, Stefanidis D. Outcome-Based Training and the Role of Simulation. 2019. p. 69–78.

[24] Velazquez-Pimentel D, Stewart E, Trockels A, et al. Global Rating Scales for the Assessment of Arthroscopic Surgical Skills: A Systematic Review. Arthroscopy. 2020;36:1156–1173.

[25] Niitsu H, Hirabayashi N, Yoshimitsu M, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. Surg Today. 2013;43:271–275.

[26] Goldenberg MG, Garbens A, Szasz P, et al. Systematic review to establish absolute standards for technical performance in surgery. Br J Surg. 2017;104:13–21.

[27] Escobar-Castillejos D, Noguez J, Neri L, et al. A Review of Simulators with Haptic Devices for Medical Training. J Med Syst. 2016;40:104.

[28] Morgan M, Aydin A, Salih A, et al. Current Status of Simulation-based Training Tools in Orthopedic Surgery: A Systematic Review. Journal of Surgical Education. 2017;74:698–716.

[29] Ruikar DD, Hegadi RS, Santosh KC. A Systematic Review on Orthopedic Simulators for Psycho-Motor Skill and Surgical Procedure Training. J Med Syst. 2018;42:168.

[30] Bartlett JD, Lawrence JE, Stewart ME, et al. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? The Bone & Joint Journal. 2018;100-B:559–565.

[31] Cecil J, Gupta A, Pirela-Cruz M. Design of VR based orthopedic simulation environments using emerging technologies. 2018 Annual IEEE International Systems Conference (SysCon). 2018. p. 1–7.

[32] Vaughan N, Dubey VN, Wainwright TW, et al. A review of virtual reality based training simulators for orthopaedic surgery. Medical Engineering & Physics. 2016;38:59–71.

[33] Verhey JT, Haglin JM, Verhey EM, et al. Virtual, augmented, and mixed reality applications in orthopedic surgery. The International Journal of Medical Robotics and Computer Assisted Surgery. 2020;16:e2067.

[34] Tay C, Khajuria A, Gupte C. Simulation training: A systematic review of simulation in arthroscopy and proposal of a new competency-based training framework. International Journal of Surgery. 2014;12:626–633.

[35] Christian MW, Griffith C, Schoonover C, et al. Construct Validation of a Novel Hip Fracture Fixation Surgical Simulator. J Am Acad Orthop Surg. 2018;26:689–697.

[36] Borgersen NJ, Naur TMH, Sørensen SMD, et al. Gathering Validity Evidence for Surgical Simulation: A Systematic Review. Annals of Surgery. 2018;267:1063–1068.

[37] Van Nortwick SS, Lendvay TS, Jensen AR, et al. Methodologies for establishing validity in surgical simulation studies. Surgery. 2010;147:622–630.

[38] Fairhurst K, Strickland A, Maddern GJ. Simulation Speak. Journal of Surgical Education. 2011;68:382–386.

[39] Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc. 2003;17:1525–1529.

[40] McDougall EM. Validation of Surgical Simulators. Journal of Endourology. 2007;21:244–247.

[41] de Ribaupierre S, Eagleson R. Editorial: Challenges for the usability of AR and VR for clinical neurosurgical procedures. Healthc Technol Lett. 2017;4:151.

[42] Downing SM, Yudkowsky R. Assessment in Health Professions Education. Routledge; 2009.

[43] Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks ? a publishing format for reproducible computational workflows. In: Loizides F, Scmidt B, editors. Positioning and Power in Academic Publishing: Players, Agents and Agendas [Internet]. IOS Press; 2016. p. 87–90. Available from: https://eprints.soton.ac.uk/403913/.

[44] Inc PT. Collaborative data science [Internet]. Montreal, QC: Plotly Technologies Inc.; 2015. Available from: https://plot.ly.

[45] team T pandas development. pandas-dev/pandas: Pandas [Internet]. Zenodo; 2020. Available from: https://doi.org/10.5281/zenodo.3509134.

[46] McKinney W. Data Structures for Statistical Computing in Python. In: Walt S van der, Millman J, editors. Proceedings of the 9th Python in Science Conference. 2010. p. 56–61.

[47] Oliphant TE. A guide to NumPy. Trelgol Publishing USA; 2006.

[48] Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Computing in Science & Engineering. 2011;13:22.

[49] Tillander B, Ledin T, Nordqvist P, et al. A virtual reality trauma simulator. Medical Teacher. 2004;26:189–191.

[50] Cobb J, Kannan V, Brust K, et al. Navigation Reduces the Learning Curve in Resurfacing Total Hip Arthroplasty. Clinical Orthopaedics and Related Research. 2007;463:90–97.

[51] Blyth P, Stott NS, Anderson IA. A simulation-based training system for hip fracture fixation for use within the hospital environment. Injury. 2007;38:1197–1203.

[52] Gofton W, Dubrowski A, Tabloie F, et al. The effect of computer navigation on trainee learning of surgical skills. J Bone Joint Surg Am. 2007;89:2819–2827.

[53] Blyth P, Stott NS, Anderson IA. Virtual reality assessment of technical skill using the Bonedoc DHS simulator. Injury. 2008;39:1127–1133.

[54] Froelich JM, Milbrandt JC, Novicoff WM, et al. Surgical Simulators and Hip Fractures: A Role in Residency Training? Journal of Surgical Education. 2011;68:298–302.

[55] Pollard TC, Khan T, Price AJ, et al. Simulated hip arthroscopy skills: learning curves with the lateral and supine patient positions: a randomized trial. Journal of Bone and Joint Surgery, American Volume [Internet]. 2012 [cited 2020 Feb 5];94. Available from: https://researchportal.bath.ac.uk/en/publications/simulated-hip-arthroscopy-skills-learning-curves-with-the-lateral.

[56] Riehl J, Widmaier J. A Simulator Model for Sacroiliac Screw Placement. Journal of Surgical Education. 2012;69:282–285.

[57] Pahuta MA, Schemitsch EH, Backstein D, et al. Virtual fracture carving improves understanding of a complex fracture: a randomized controlled study. J Bone Joint Surg Am. 2012;94:e182.

[58] Nousiainen MT, Omoto DM, Zingg PO, et al. Training femoral neck screw insertion skills to surgical trainees: computer-assisted surgery versus conventional fluoroscopic technique. Journal of Orthopaedic Trauma. 2013;27:87–92.

[59] Cassar-Gheiti AJ, Dodds MK, Byrne DP, et al. Preliminary study of the feasibility and accuracy of percutaneous peri-acetabular screw insertion in a porcine model. Injury. 2013;44:178–182.

[60] Rambani R, Viant W, Ward J, et al. Computer-Assisted Orthopedic Training System for Fracture Fixation. Journal of Surgical Education. 2013;70:304–308.

[61] Pedersen P, Palm H, Ringsted C, et al. Virtual-reality simulation to assess performance in hip fracture surgery. Acta Orthopaedica. 2014;85:403–407.

[62] Akhtar K, Sugand K, Sperrin M, et al. Training safer orthopedic surgeons. Construct validation of a virtual-reality simulator for hip fracture surgery. Acta Orthopaedica. 2015;86:616–621.

[63] Kho JY, Johns BD, Thomas GW, et al. A Hybrid Reality Radiation-Free Simulator for Teaching Wire Navigation Skills. J Orthop Trauma. 2015;29:e385-390.

[64] Sugand K, Akhtar K, Khatri C, et al. Training effect of a virtual reality haptics-enabled dynamic hip screw simulator. Acta Orthopaedica. 2015;86:695–701.

[65] Sugand K, Mawkin M, Gupte C. Training effect of using Touch Surgery™ for intramedullary femoral nailing. Injury. 2016;47:448–452.

[66] Dwyer T, Wadey V, Archibald D, et al. Cognitive and Psychomotor Entrustable Professional Activities: Can Simulators Help Assess Competency in Trainees? Clin Orthop Relat Res. 2016;474:926–934.

[67] Khanduja V, Lawrence JE, Audenaert E. Testing the Construct Validity of a Virtual Reality Hip Arthroscopy Simulator. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2017;33:566–571.

[68] Phillips L, Cheung JJH, Whelan DB, et al. Validation of a Dry Model for Assessing the Performance of Arthroscopic Hip Labral Repair. Am J Sports Med. 2017;45:2125–2130.

[69] Condino S, Turini G, Parchi PD, et al. How to Build a Patient-Specific Hybrid Simulator for Orthopaedic Open Surgery: Benefits and Limits of Mixed-Reality Using the Microsoft HoloLens [Internet]. Journal of Healthcare Engineering. 2018 [cited 2020 Jan 21]. Available from: https://new.hindawi.com/journals/jhe/2018/5435097/.

[70] Erturan G, Alvand A, Judge A, et al. Prior Generic Arthroscopic Volume Correlates with Hip Arthroscopic Proficiency: A Simulator Study. The Journal of Bone and Joint Surgery [Internet]. 2018 [cited 2020 Feb 1];100. Available from: insights.ovid.com.

[71] Sugand K, Wescott RA, Carrington R, et al. Teaching basic trauma: validating FluoroSim, a digital fluoroscopic simulator for guide-wire insertion in hip surgery. Acta Orthopaedica. 2018;89:380–385.

[72] Vaghela KR, Lee J, Akhtar K. Performance on a Virtual Reality DHS Simulator Correlates with Performance in the Operating Theatre. Surg Technol Int. 2018;33.

[73] Christian MW, Griffith C, Schoonover C, et al. Construct Validation of a Novel Hip Fracture Fixation Surgical Simulator. J Am Acad Orthop Surg. 2018;26:689–697.

[74] Logishetty K, Western L, Morgan R, et al. Can an Augmented Reality Headset Improve Accuracy of Acetabular Cup Orientation in Simulated THA? A Randomized Trial. Clinical Orthopaedics and Related Research. 2018;Publish Ahead of Print:1.

[75] Aguilera-Canon MC, Wainwright T, Yang X, et al. Mixed Reality-Based Simulator for Training on Imageless Navigation Skills in Total Hip Replacement Procedures. In: El Rhalibi A, Pan Z, Jin H, et al., editors. E-Learning and Games. Cham: Springer International Publishing; 2019. p. 30–34.

[76] Gallagher K, Bahadori S, Antonis J, et al. Validation of the Hip Arthroscopy Module of the VirtaMed Virtual Reality Arthroscopy Trainer. Surgical technology international. 2019;34:430–436.

[77] FundamentalVR. Fundamental Surgery conducts validation study with BOTA: British Orthopeadic Trainee Association [Internet]. Fundamental Surgery. 2019 [cited 2020 Feb 4]. Available from: https://www.fundamentalsurgery.com/british-orthopaedic-trainees-association/.

[78] Panariello D, Caporaso T, Grazioso S, et al. Using the KUKA LBR iiwa Robot as Haptic Device for Virtual Reality Training of Hip Replacement Surgery. 2019 Third IEEE International Conference on Robotic Computing (IRC). 2019. p. 449–450.

[79] Weik D, Lorenz M, Knopp S, et al. Integrating Tactile Feedback in an Acetabular Reamer for Surgical VR-Training. 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2019. p. 1227–1228.

[80] Bauer DE, Wieser K, Aichmair A, et al. Validation of a Virtual Reality-Based Hip Arthroscopy Simulator. Arthroscopy: The Journal of Arthroscopic & Related Surgery. 2019;35:789–795.

[81] Gustafsson A, Pedersen P, Rømer TB, et al. Hip-fracture osteosynthesis training: exploring learning curves and setting proficiency standards. Acta Orthopaedica. 2019;90:348–353.

[82] Sugand K, Wescott RA, Carrington R, et al. Training and Transfer Effect of FluoroSim, an Augmented Reality Fluoroscopic Simulator for Dynamic Hip Screw Guidewire Insertion: A Single-Blinded Randomized Controlled Trial. J Bone Joint Surg Am. 2019;101:e88.

[83] Arroyo-Berezowsky C, Jorba-Elguero P, Altamirano-Cruz MA, et al. Usefulness of immersive virtual reality simulation during femoral nail application in an orthopedic fracture skills course. Journal of Musculoskeletal Surgery and Research. 2019;3:326.

[84] Bartlett JD, Lawrence JE, Khanduja V. Virtual reality hip arthroscopy simulator demonstrates sufficient face validity. Knee Surg Sports Traumatol Arthrosc. 2019;27:3162–3167.

[85] Weber A, Domes C, Christian M, et al. Effect of Training Modules on Hip Fracture Surgical Skills Simulation Performance: Early Validation of the AAOS/OTA Simulator. The Journal of Bone and Joint Surgery. 2019;101:2051–2060.

[86] Cychosz C, Khazi ZM, Karam M, et al. Validation of a novel hip arthroscopy simulator: establishing construct validity. J Hip Preserv Surg. 2019;6:385–389.

[87] Logishetty K, Gofton WT, Rudran B, et al. A Multicenter Randomized Controlled Trial Evaluating the Effectiveness of Cognitive Training for Anterior Approach Total Hip Arthroplasty. J Bone Joint Surg Am. 2020;102:e7.

[88] Logishetty K, Gofton WT, Rudran B, et al. Fully Immersive Virtual Reality for Total Hip Arthroplasty: Objective Measurement of Skills and Transfer of Visuospatial Performance After a Competency-Based Simulation Curriculum. JBJS [Internet]. 2020 [cited 2020 Jan 21];Latest Articles. Available from: https://journals.lww.com/jbjsjournal/Abstract/latest/Fully_Immersive_Virtual_Reality_for_Total_Hip.99864.aspx.

[89] Bartlett JD, Lawrence JE, Yan M, et al. The learning curves of a validated virtual reality hip arthroscopy simulator. Arch Orthop Trauma Surg [Internet]. 2020 [cited 2020 Feb 2]; Available from: https://doi.org/10.1007/s00402-020-03352-3.

[90] Rölfing JD, Jensen RD, Paltved C. HipSim - hip fracture surgery simulation utilizing the Learning Curve-Cumulative Summation test (LC-CUSUM). Acta Orthopaedica. 2020;0:1–6.

[91] Wang C, Ouyang Y, Liu H, et al. Surgery simulation teaching based on real reconstruction aid versus traditional surgical live teaching in the acquisition of an adult total hip arthroplasty surgical technique for developmental dysplasia of the hip: a randomized comparative study. BMC Med Educ. 2020;20:228.

[92] Geswell M, Sinha N, Mandel M, et al. Improving resident education through unstable chicken hips: a novel way to teach an infant hip examination. Journal of Pediatric Orthopaedics B [Internet]. 2020 [cited 2020 Aug 4];Publish Ahead of Print. Available from: https://journals.lww.com/jpo-b/Abstract/9000/Improving_resident_education_through_unstable.98791.aspx.

[93] Rangarajan K, Davis H, Pucher PH. Systematic Review of Virtual Haptics in Surgical Simulation: A Valid Educational Tool? Journal of Surgical Education. 2020;77:337–347.

[94] Carter FJ, Schijven MP, Aggarwal R, et al. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc. 2005;19:1523–1532.

[95] Panariello D, Caporaso T, Grazioso S, et al. Using the KUKA LBR iiwa Robot as Haptic Device for Virtual Reality Training of Hip Replacement Surgery. 2019 Third IEEE International Conference on Robotic Computing (IRC). 2019. p. 449–450.

[96] Blackhall VI, Cleland J, Wilson P, et al. Barriers and facilitators to deliberate practice using take-home laparoscopic simulators. Surg Endosc. 2019;33:2951–2959.

[97] Nyström M, Andersson R, Holmqvist K, et al. The influence of calibration method and eye physiology on eyetracking data quality. Behav Res. 2013;45:272–288.

[98] McBride TJ, Vandayar N, Nixon KJ. A Comparison of Skin Detection Algorithms for Hand Gesture Recognition. 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA). 2019. p. 211–216.

[99] Cook DA. Much ado about differences: why expert-novice comparisons add little to the validity argument. Adv Health Sci Educ Theory Pract. 2015;20:829–834.

[100] Gustafsson A, Pedersen P, Rømer TB, et al. Hip-fracture osteosynthesis training: exploring learning curves and setting proficiency standards. Acta Orthopaedica. 2019;90:348–353.

[101] Bosse HM, Mohr J, Buss B, et al. The benefit of repetitive skills training and frequency of expert feedback in the early acquisition of procedural skills. BMC Med Educ. 2015;15:22.

[102] Khatri C, Sugand K, Anjum S, et al. Does Video Gaming Affect Orthopaedic Skills Acquisition? A Prospective Cohort-Study. PLoS One [Internet]. 2014 [cited 2020 Feb 5];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4198251/.

[103] Sadideen H, Alvand A, Saadeddin M, et al. Surgical experts: Born or made? International Journal of Surgery. 2013;11:773–778.

[104] Stefanidis D, Jr JRK, Sweet RM, editors. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties [Internet]. Springer International Publishing; 2019 [cited 2020 Apr 19]. Available from: https://www.springer.com/gp/book/9783319982755.

[105] Korndorffer JR. Principles of Validity. In: Stefanidis D, Korndorffer Jr. JR, Sweet R, editors. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties [Internet]. Cham: Springer International Publishing; 2019 [cited 2020 Apr 19]. p. 37–39. Available from: https://doi.org/10.1007/978-3-319-98276-2_4.

[106] Vandenberg SG, Kuse AR. Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. Percept Mot Skills. 1978;47:599–604.

[107] Gardner A, Lau JN, Kim S. A Taxonomy Guide for Surgical Simulation. In: Stefanidis D, Korndorffer Jr. JR, Sweet R, editors. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties [Internet]. Cham: Springer International Publishing; 2019 [cited 2019 Jun 13]. p. 25–36. Available from: https://doi.org/10.1007/978-3-319-98276-2_3.

[108] Julian DL, Smith RD, Tanaka ADS, et al. Design methodology for a simulator of a robotic surgical system. J Robotic Surg [Internet]. 2018 [cited 2019 Jun 13]; Available from: https://doi.org/10.1007/s11701-018-0897-z.

[109] Gupta N, Agrawal H. COVID 19 and Surgical Education: Time for Innovations. Indian J Surg. 2020;1–2.

[110] Das De S, Puhaindran ME, Sechachalam S, et al. Sustaining a national surgical training programme during the COVID-19 pandemic. Bone & Joint Open. 2020;1:98–102.

[111] Blanco-Colino R, Soares AS, Kuiper SZ, et al. Surgical Training During and After COVID-19: A Joint Trainee and Trainers Manifesto. Annals of Surgery. 2020;272:e24.

[112] English W, Vulliamy P, Banerjee S, et al. Surgical training during the COVID-19 pandemic – the cloud with a silver lining? BJS (British Journal of Surgery). 2020;107:e343–e344.

[113] McKechnie T, Levin M, Zhou K, et al. Virtual Surgical Training During COVID-19. Ann Surg. 2020;272:e153–e154.

[114] Stambough JB, Curtin BM, Gililland JM, et al. The Past, Present, and Future of Orthopedic Education: Lessons Learned From the COVID-19 Pandemic. J Arthroplasty. 2020;35:S60–S64.

[115] Ali A, Subhi Y, Ringsted C, et al. Gender differences in the acquisition of surgical skills: a systematic review. Surg Endosc. 2015;29:3065–3073.