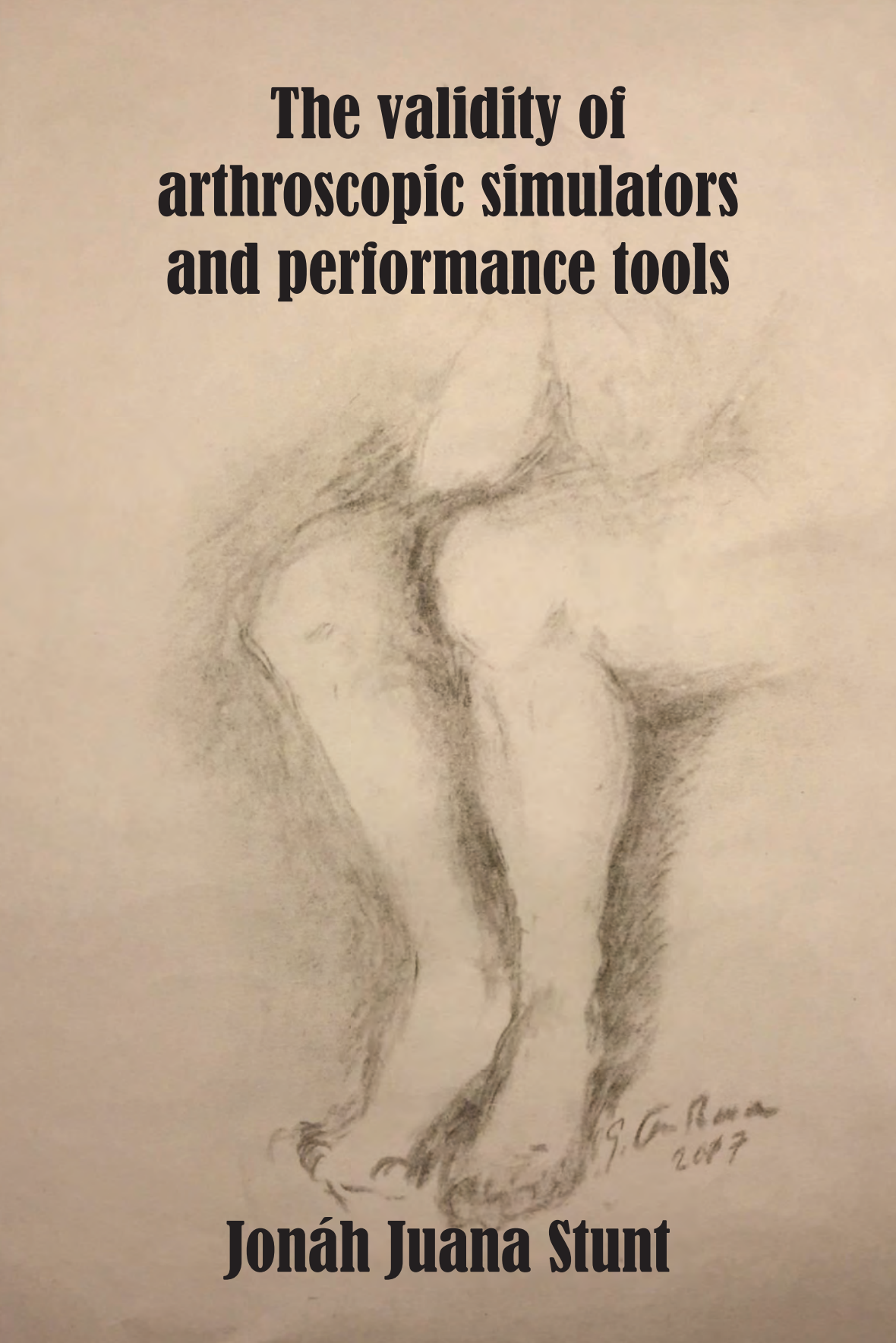
The validity of arthroscopic simulators and performance tools

Stunt, J.J.

Publication date
2017
Document Version
Final published version
License
Other

The validity of arthroscopic simulators and performance tools



Jonáh Juana Stunt

The validity of arthroscopic simulators and performance tools

Jonáh Juana Stunt

PhD Thesis of
Jonáh Stunt, MSc
PhD Research Fellow
Department of Orthopaedic Surgery
Academic Medical Center
University of Amsterdam

Thesis University of Amsterdam
Cover design: Maria Giovanna Aymerich

This thesis was supported by:

AMC Graduate School, European Society of Sports Traumatology Knee Surgery and Arthroscopy (ESSKA), Marti Koning Eckhart Stichting, Medishield, Simendo, Slotervaart ziekenhuis, Smith & Nephew NL, Stanford University School of Medicine, Tergooi ziekenhuis, TU Delft and Virtamed, which is gratefully acknowledged.

The validity of arthroscopic simulators and performance tools

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

Prof. dr. ir. K.I.J. Maex

ten overstaan van een door het college voor promoties ingestelde commissie,

in het openbaar te verdedigen in de Agnietenkapel

op woensdag 18 januari 2017, te 10.00 uur

door Jonáh Juana Stunt

geboren te Amsterdam

PROMOTIE COMMISSIE

Promotores
Prof. dr. G.M.M.J. Kerkhoffs, UvA
Prof. dr. C.N. van Dijk, UvA

Co-promotor
Dr. ir. G.J.M Tuijthof , Zuyd Hogeschool

Overige leden
Prof. dr. J. Dankelman, TU Delft
Prof. dr. R.L. Diercks, RUG
Prof. dr. M. Maas, UvA
Prof. dr. M.J.H.M. Lombaerts, UvA
Dr. M.C. Obdeijn, UvA
Prof. dr. M.P. Schijven, UvA

Faculteit der Geneeskunde

Voor papa en mama

Table of contents

CHAPTER I

INTRODUCTION

Arthroscopies are frequently performed procedures in orthopedic surgery and are presently the gold standard treatment for intra-articular lesions. The number of annually performed arthroscopies continues to grow[1, 2]. Arthroscopic techniques will evolve for the greater part due to the upcoming cartilage tissue replacement methods and the demands from high performance athletes who sustain sports injuries[3, 4]. As acquiring arthroscopic skills is challenging for orthopedic residents, effective and efficient training is a prerequisite before performing arthroscopic procedures in the operating room on real patients. Orthopedic residency programs have remained practically unchanged for a long time, but with the increasing restrictions on work hours and more focus on patient safety, a paradigm shift in orthopedic residency training is inevitable[5-7]. Hence, the use of simulation for arthroscopic surgery could eventually become an important alternative method to acquire arthroscopic psychomotor skills safely, efficiently and effectively[8-12].

In this thesis, the focus lies on arthroscopic skills training, more specifically, on the development and validity of simulation environments for training, as this is key to acceptance and implementation in residency programs. As an introduction, the traditional arthroscopic training and its advantages and disadvantages, will be reviewed first. Thereafter, an overview of alternative simulators will be provided. Thirdly, as monitoring and assessing performance is essential to offer adequate and effective training, the available performance tools and metrics will be discussed. For the latter, performance metrics and complementary thresholds need to be determined and validated to indicate the margins for safe and efficient arthroscopic skills and provide feedback during training.

Arthroscopy and traditional skills training

Arthroscopy originated in Japan in the 1930s[13]. The technique eventually became popular and has been increasingly used for orthopedic interventions, due to advantages compared to open surgery: lower morbidity, higher diagnostic accuracy and greater therapeutic efficacy[14, 15]. Arthroscopic procedures are performed on joints, using arthroscopic instruments that are inserted in the joint through a couple of small incisions. An arthroscopic system consists of an arthroscope, a light source- and cable, a camera system and video recorder, and arthroscopic instruments. The basic arthroscopic instrument set consist of a probe, basket and grasping forceps, arthroscopic scissors and knives (mechanical instruments) and a shaver (motorized instrument). Arthroscopies can be performed on every joint, but knee arthroscopy is one of the most frequently performed procedures with 4 million times a year, and is applied for a range of pathologies, such as meniscal tears, cartilage flaps lesions, loose body removal, anterior and posterior cruciate ligament reconstruction, meniscal transplantation, and synovectomy[2].

Arthroscopic skills are complex, due to an indirect view on the pathologic area through an arthroscope, disturbed tactile feedback, complex hand-eye-coordination because of the three-dimensional surgical environment shown on a two-dimensional display at a 30 degree angled view, and manual bilateral handling of different surgical procedures in the narrow joint space through two or three small entry points[15-17] (Figure 1). This demands a completely different skills-set compared with open orthopedic surgery, which present direct vision and haptic feedback and enable digital palpation of tissue. Moreover, many of the open surgical techniques use skills similar to the ones used in daily life, from which more advanced surgical techniques are intuitively developed[15]. Thus,

arthroscopic procedures ask for a high level of psychomotor skills, that lead to a long learning curve, and need to be sufficiently trained to ensure patient safety in the operating room.

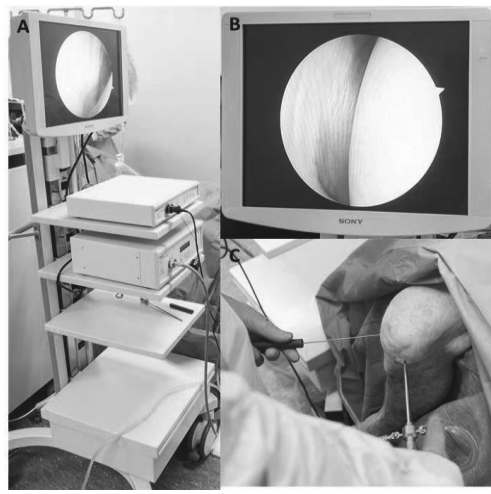


Figure 1: A. An arthroscopy tower with a light source, a camera system and a monitor **B.** monitor displaying the inside of the knee joint **C.** surgeon inserting the arthroscope and probe into the joint space through two entry points (portals)

Arthroscopic training usually begins during orthopaedic residency. Traditionally, written material, lectures and seminars and observational learning are part of arthroscopic training[15]. Moreover, cadaver material has since long been used to train arthroscopic skills. Cadaver training is the gold standard and most arthroscopic residency programs still start training of arthroscopic skills on cadavers (Figure 2)[13, 18]. Although bovine and porcine animal models present a similar gross and arthroscopic anatomy as the human knee joint, human cadavers are preferred both by teaching staff and residents to practice arthroscopic skills, as they enable to study exact human anatomy. Cadavers, human cadavers specifically, are the most realistic mode of simulation for arthroscopic training in terms of tissue appearance, anatomy and sensory feedback[19-22]. Moreover, they are easy to handle, assembly and store[13]. A drawback, is that cadaver

courses are not fairly costly and offer insufficient exposure to truly make the required hours to acquire arthroscopic skills. Finally, training consists for the majority of a one-to-one apprentice model in the OR, where the resident performs procedures on patients under supervision of an experienced surgeon.

There are some limitations of the traditional educational paradigm that undermine the efficiency and benefits of the approach[15,] economy induced recent transformations in surgical training programs have resulted in reduced training time, work hour restriction and decreased case numbers[10, 23, 24]. Residents are expected to perform arthroscopic procedures in the operating theater shortly after they have started their training. But as the quantity and quality of residents' surgical training is compromised, orthopedic residents feel less prepared to perform arthroscopic procedures compared with open procedures[25]. This is associated with reduced quality of patient care, less patient safety, and increased damage of healthy tissues[7, 25-27]. Moreover, arthroscopic training in the operating room is relatively inefficient as it significantly increases the duration of the surgical procedure[15]. There are also a number of objections to the use of cadavers: they are expensive, difficult to procure, they have a lack of feedback and tissue may not always be natural or intact[13, 28]. Fresh cadavers have limited time of use and are not reusable. Moreover, human cadaver teaching is restricted by ethic rules[18, 28]. Thus, the traditional approach seems neither the most efficient or effective form of training, nor does it meet patient safety standards sufficiently. Therefore, the traditional arthroscopic approach is difficult to maintain[10, 12].

Figure 2: Student practicing arthroscopic skills on a cadaver knee

Arthroscopic simulation

Recent changes in the healthcare system have enforced a growing need for alternative training methods to train arthroscopic skills outside the operating theatre[29]. The use of simulators is such an alternative. Medical simulation started in ancient times, with the use of sheep lungs and livers as primitive forms of physical models[30]. From then, scientists never stopped working on the development of simulation, until late 20th century academic centers and research organizations began to develop more realistic models of the human body. In the 1990s, inspired by simulation in the military and aviation industry, and facilitated by advances in computer science and other technologies, basic (virtual reality) training simulators were developed and evolved into advanced and sophisticated simulators[30-32] (Figure 3). Today, three categories of arthroscopic simulators exist according to their resemblance to reality: low-fidelity, medium-fidelity and high-fidelity simulators[16].

Low-fidelity simulators are often static, without realism or situational context, and are appropriate for teaching the basics of technical

medical skills. They are represented by box trainers and physical models. Box trainers do not resemble a human joint, but are an actual box (Figure 4). Physical models, such as anatomic bench models, are artificial models that represent the bones and joints of the body (Figure 5)[33]. Box trainers and physical models can be used to train basic psychomotor skills including eye-hand coordination, precise manipulation and bimanual tasks [13]. Both models can provide sensory feedback and can be equipped with sensors to objectively monitor performance based on motion and force metrics[17]. Medium-fidelity simulators have more resemblance with reality and can be used for training of specific, increasingly complex competencies. High-fidelity simulators are realistic and dynamic, and combine a body (part) model with computers[34]. Some anatomic bench models can be characterized as medium- or high-fidelity simulators, depending on the degree to which they resemble a human joint. High-fidelity simulators are mainly represented by virtual reality simulators combined with a synthetic joint model (Figure 6 and 7). VR arthroscopic simulators comprise a computer and screen (the virtual world) and instruments that are similar to those used in the operating theatre. The elements of the joint's anatomy and instruments are displayed in a computer-simulated environment. The instruments can be used to manipulate virtual tissues and organs. There may also be visual feedback through the visible deformity of tissues or force feedback through a haptic device[17]. The synthetic joint model can be manipulated through predefined portals into which instruments can be inserted. Key elements to VR simulators are the visual graphics, haptic feedback with tissue deformation and feedback on performance[13]. For an overview of currently available arthroscopic simulators, see Table 1.

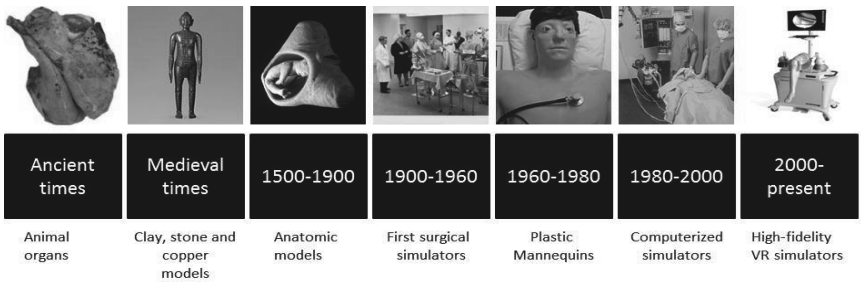


Figure 3: Timeline of simulator development

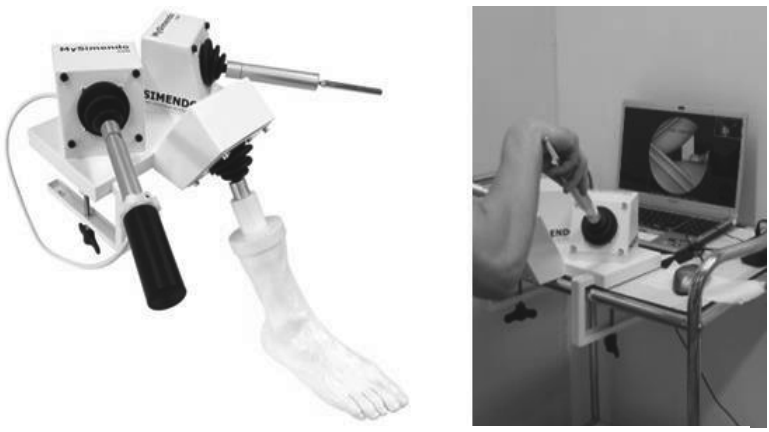


Figure 4: Simendo Pro arthroscopy for knee arthroscopy with virtual display

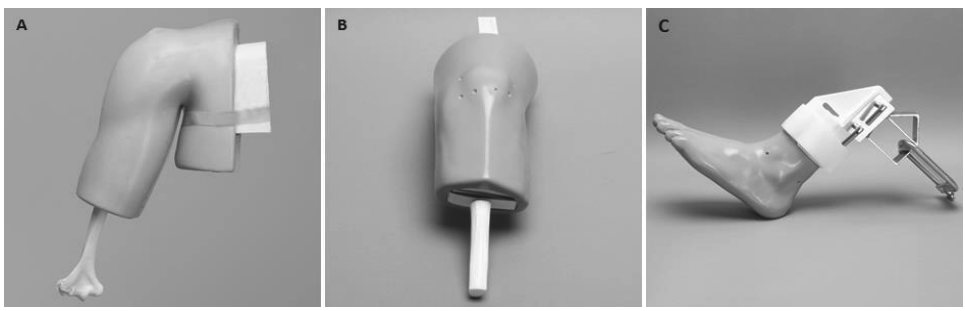


Figure 5: A. Sawbones Arthroscopy Shoulder with Replaceable Components (© Sawbones Europe AB, 2014). **B**. ACL knee trainer (© Sawbones Europe AB, 2014). **C**. Sawbones Arthroscopy Foot and Ankle (© Sawbones Europe AB, 2014).
Reprinted with permission from www.sawbones.com)

Table 1: An overview of currently available simulators to practice arthroscopic skills, from low fidelity to high fidelity

Currently available arthroscopic simulators		
Low fidelity	**Box trainers**	FAST Arthroscopy Training Workstation (www.sawbones.com)
		Box Trainer for Arthroscopic Knot Tying (www.rcseng.ac.uk)
		Arthroscopy training model for junior trainees (www.rcseng.ac.uk)
		SIMENDO Arthroscopy (www.simendo.eu)
	Anatomic bench models	Sawbones knee, shoulder and wrist joint bench models (dry and wet) (www.sawbones.com)
		Adam-Rouilly knee and shoulder joint bench models (www.adamrouilly.co.uk)
		Hillway Surgical Knee and shoulder joint bench models (www.surgimodels.com)
		The CLA knee, shoulder and wrist joint joint bench models (www. coburger-lehrmittelanstalt.de)
		Beijing Yimo Shoulder Joint (www.chinamedevice.com)
		Wrist bench model (Academic Medical Centre in Amsterdam and Delft University of Technology, not
	Anatomic bench models with sensors	Alex Shoulder Professor(www.sawbones.com)
		Knee joint bench model(www.sawbones.com)
		ForceTRAP v2 (www.medishielddelft.com)
	VR simulators with synthetic joint model	Knee Arthroscopy Surgical Trainer: ArthroSim TM (www.toltech.net)
		ARTHRO Mentor TM for knee and shoulder (www.simbionix.com)
		VirtaMed ArthroS TM for Knee and Shoulder (www.virtamed.com)
High fidelity	**Hybrid Models**	Practice Arthroscopic Surgical Skills for Perfect Operative Real-Life Treatment (PASSPORT)
		Knee Arthroscopy Simulator[71]

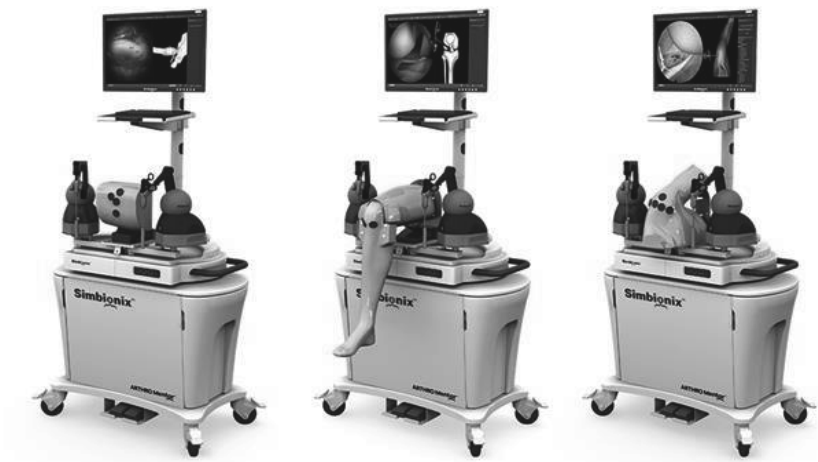


Figure 6: ARTHRO Mentor™ VR simulators for hip, knee and shoulder arthroscopy (Reprinted with permission from www.simbionix.com**)**

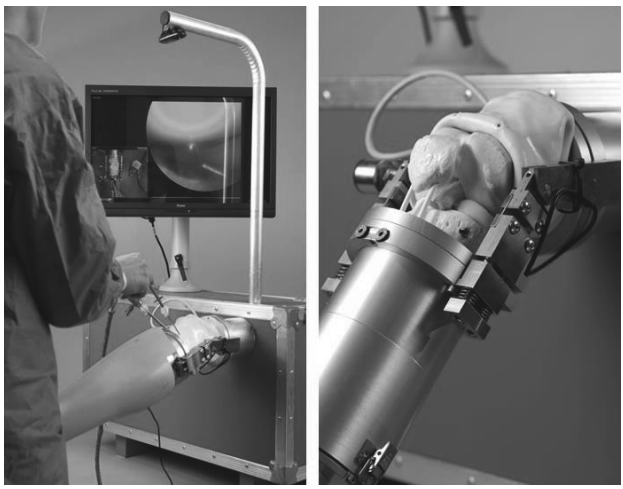


Figure 7: The PASSPORT (Practice Arthroscopic Surgical Skills for Perfect Operative Real-life Treatment) for knee arthroscopy

Performance tracking

The effect of simulator training on arthroscopic skills needs to be evaluated by means of performance tracking. It has been demonstrated that performance tracking is only beneficial if feedback on performance is given to the trainee: direct objective feedback stimulates and supports the learning process and improves the trainees' individual skills[35-38]. In order to give adequate feedback, thresholds for performance metrics need to be set and categories for evaluation criteria and different levels of competence are necessary to be formulated and validated.

Performance Metrics and Thresholds

Performance tracking requires objective metrics able to reflect ones learning capability and learning curve. A range of metrics for performance efficiency and performance safety is available to objectively monitor arthroscopic performance (Table 2)[17]. Metrics are registered with sensors built in a simulator[17]. To develop metrics that are representative for a certain task and skill training an important step is to derive thresholds. Firstly, a comparison of objective measures of skills in the surgical environment must be associated with simulator metrics[39]. Necessary measurements need to be performed to represent the properties and physiological processes accurately, taking into account the variation in tissue material properties amongst the human population. For example, to enable correct training of force application on meniscal material, deformation zones for elastic and plastic deformation of human material need to be determined. Menisci samples are collected and surgeons have to perform several tasks to measure applied forces in different directions. Average applied forces, peak forces and overall forces overtime must be

measured. Variation in forces between the menisci and between the surgeons are compared, and the maximum applied forces is compared with the theoretical maximum force; the minimum force that will damage meniscal tissue when exerting force with a probe[40]. When threshold for performance metrics are determined, metrics and thresholds have to be integrated into the simulator[17, 32]. This process supports valid simulator training curricula that can offer performance tracking, can correct and enhance performance, and can provide exercises that discriminate between levels of experience[32].

Table 2: Metrics for performance efficiency and safety

Performance Efficiency	Performance Safety
Task repetition	Collisions
Task error	Out of view time
Task time	Tip-to-tip distance
Idle time	Motion speed
Path length	Motion smoothness
Economy of movement	Force magnitude
Depth perception	Force direction
Volume of motion	Force area
	Volume of force

Evaluating and monitoring arthroscopic performance

Performance and safety metrics provide feedback for part-task skills, but to adequately define proficiency, a more holistic, expert judgment is required as well. In the traditional apprenticeship model, the trainee is assessed by a supervisor who rates the trainee's skills globally without using specific criteria[41]. This type of rating is largely subjective, and

possesses poor test-retest and inter-observer reliability, as the judgment of trainee's skills can vary significantly among supervisors[42, 43]. Subjective assessment devalues the evaluation of individual proficiency and hinders adequate differentiation between novice, intermediate, and expert levels in surgical performance. Set criteria against which skills can be evaluated make assessment of trainees by supervisors more objective, valid, and reliable. Global rating scales (GRS) are eligible tools to objectively monitor complex task and assess performance in a more holistic approach.

A GRS provides a qualitative outcome measure of technical skills necessary for performing a procedure, such as instrument handling, flow of operation, knowledge of the specific procedure, autonomy, efficiency, and quality of the operative result[42]. Global rating scales consist of several domains that describe clear evaluation criteria and various levels of competence. Each domain needs to be scored on a 5-point Likert scale enabling uniform assessment[42]. As the concept of objective structured assessment for technical skills has been widely accepted by various surgical fields[16, 44-48], many operation-specific global rating scales have been developed[41, 49-53]. However, it is relatively new to arthroscopic training[54]. Recently, various GRS have been developed specifically for monitoring and training of arthroscopic skills (see Table 3), presenting similar demands with respect to what a resident should be able to demonstrate in the operating theatre and the required level to qualify as novice, competent or expert[17].

Table 3: Global Rating Scales for arthroscopic skills performance
• Orthopedic Competence Assessment Project (OCAP) [10]
• Basic Arthroscopic Knee Skill Scoring System (BAKSSS)[42]
• Arthroscopic Skills Assessment (ASA) [41]
• Objective Assessment of Arthroscopic Skills (OAAS)[54]
• Arthroscopic Surgery Skill Evaluation Tool (ASSET) [55]

Research gaps in validity of simulators and effect of simulator training of arthroscopic skills

The development of simulators and simulator programs for arthroscopic surgery has been lagging behind compared to other surgical disciplines[13], arthroscopic simulators and training programs have become available relatively recently[5, 18, 29, 56-58].

However, these simulators have drawbacks: the main drawback of most physical models is their lack of realism or low fidelity, and their limited ability to model physiologic processes, pathologies and anatomic variability[13, 56, 59, 60]. Secondly, they need continuous repair and maintenance[13]. Moreover, although box trainers and bench models can be combined with sensors, most models lack an adequate graphical user interface, automatic data processing and performance feedback[40, 61]. Besides, even if sensors are provided, these models still lack performance registration and skill progression monitoring accordingly, which implies that supervisor-independent learning cannot be provided[11]. These shortcomings undermine the widespread integration of the models in arthroscopic training curricula[13, 56].

High-fidelity VR simulators overcome many of the limitations of physical models[18]. Performance metrics can be documented and used for feedback and training progression. Moreover, VR simulators are less resource- and staff intensive[62] as they are more appropriate to supervisor-

independent learning[28]. However, controversy on the effectiveness of (high-fidelity) arthroscopic simulators exists and convincing evidence of the validity of these tools is required. There have been various studies on different levels of validity and reliability of simulators. Several simulators present sufficient realism[63-67] and are able to discriminate between novice and expert surgeons[10, 28, 63, 65-71]. Of great concern is the fact that it has not been clarified yet whether arthroscopic simulators are able to discriminate between groups that are more closely associated with each other in terms of arthroscopic experience and skills proficiency. Moreover, there is limited evidence on the contribution of simulator training to proficiency in arthroscopic skills, and the transfer of skills from simulator training to the operating theatre has been scarcely studied.

There are three other concerns with regard to studies on knee simulator validity that obstruct simulators to become accepted as useful tools in arthroscopic education[26].

- Heterogeneity in the literature exists on validation studies, impeding comparison between simulators;
- There is a range of limitations and flaws on the methodology used in the studies[12];
- Used assessment methods are not standardized or validated[12].

As a consequence, most arthroscopic training programs do not use simulators routinely[9], even though in other disciplines simulator technology has proven their benefit[30].

Research gaps in validity of performance monitoring tools

So far, the focus has mainly been on performance metrics, in particular task completion. However, results on studies on safety metrics for tissue manipulation demonstrate that this is an equally important performance category[71, 72]. Excessive stressing forces on a joint can damage intra-articular structures, and thus safe execution of joint stressing is important. Development of such safety metrics is in its infancy for arthroscopic skills training.

With the ever-increasing number of yearly performed arthroscopies, valid and reliable scoring methods are vital to evaluate individual proficiency and to distinguish between novice, intermediate, and expert levels in diagnostic knee arthroscopy. More research on the validity of rating scales is needed to investigate whether they allow objective evaluation of arthroscopic skills proficiency and development, and learning curve assessment.

Thesis aims

In view of the shortcomings and missing evidence as stated, the aim of this thesis is to:

1. Demonstrate simulator validity using standardization protocols and improved study designs and methodology.
2. Validate safety metrics and complementary thresholds for safe and non-damaging tissue handling.
3. Validate global rating scales for objective measurement and monitoring of arthroscopic performance and indicate guidelines for implementation.

Thesis outline

Validation of training and assessment tools is becoming increasingly important in arthroscopic surgery[73]. As the concept of validity is running through this thesis as a common theme, Chapter 2 will be dedicated to the principles of validity and its different levels and types.

In Chapter 3 an overview of all commercially available medical simulators is provided and their validity evidence determined, to inform hospitals and medical educators who wish to implement simulation training in their curriculum about the suitability of various types of simulators.

Methodological flaws and limitations have generated controversy about the validity of devices for knee arthroscopic training[12]. In an attempt to overcome a number of these limitations, face and construct validity of two high-fidelity knee arthroscopy simulators will be assessed with a standardized and carefully developed study protocol in Chapter 4 and 5. As the lack of randomized controlled trials and transfer validity evidence is of particular concern, a protocol is set up in order to cater for this shortcoming, which will be elaborated in Chapter 6.

As safe loading levels for joint stressing have not yet been set and integrated in simulators, joint stressing thresholds need to be validated. Chapter 7 focuses on force magnitude thresholds for safe ligament loading to prevent tissue damaging when performing tissue manipulation.

Global Rating Scales for monitoring arthroscopic skills have not been investigated for their suitability to track individual arthroscopic training progress. Therefore, in Chapter 8, the validity and utility to assess performance and monitor individual learning of two of these GRSs will be examined.

References

[1] Cullen KA, Hall MJ, Golosinskiy A: Ambulatory surgery in the United States. Natl Health Stat Report 2009, 11:25-53.

[2] Kim S, Bosque J, Meehan JP, Jamali A, Marder R: Increase in outpatient knee arthroscopy in the United States: a comparison of National Surveys of Ambulatory Surgery, 1996 and 2006. J Bone Joint Surg 2011, 93:994-1000.

[3] Byrd T: Hip Arthroscopy in The Athlete. Am J Sports Phys Ther 2007, 2:217-230.

[4] Duretti FT, Goldfarb CA: Sports injuries of the wrist. Curr Rev Musculoskelet Med 2013, 6:35-40.

[5] Aim F, Lonjon G, Hannouche D, Nizard R: Effectiveness of Virtual Reality Training in Orthopaedic Surgery. Arthroscopy 2015, 32:224-232.

[6] Alvand A, Auplish S, Gill H, Rees J: Innate arthroscopic skills in medical students and variation in learning curves. J Bone Joint Surg Am 2011, 93:e115(1-9).

[7] Peabody T, Nestler S, Marx C, Pellegrini V: Resident duty-hour restrictions-who are we protecting? AOA critical issues. J Bone Joint Surg Am 2012, 94:E131: 1-7.

[8] Cannon WD, Garret Jr. WE, Hunter RE, Sweeney HJ, Eckhoff DG, Nicandri GT, Hutchinson MR, Johnson DD, Bisson LJ, Bedi A, Hill JA, Koh JL, Reinig KD: Improving Residency Training in Arthroscopic Knee Surgery with Use of a Virtual-Reality Simulator: A Randomized Blinded Study. J Bone Joint Surg Am 2014, 96:1798-1806.

[9] Gomoll AH, Pappas G, Forsythe B, Warner JJ: Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. Am J Sports Med 2008, 36:1139-1142.

[10] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Am 2008, 90:494-499.

[11] Jacobsen ME, Anderson MJ, Hansen CO, Konge L: Testing basic competency in knee arthroscopy using a virtual reality simulator: exploring validity and reliability. J Bone Joint Surg Am 2015, 97:775-781.

[12] Modi CS, Morris G, Mukherjee R: Computer-simulation training for knee and shoulder arthroscopic surgery. Arthroscopy 2010, 26:832-840.

[13] Madan SS, Pai DR: Role of Simulation in Arthroscopy Training. Simul Healthc 2014, 9:127-135.

[14] Garrett Jr. WE, Swiontkowski M, Weinstein JN, Callaghan J, Rosier RN, Berry DJ, Harrast J, Derosa GP: American Board of Orthopaedic Surgery

Practice of the Orthopaedic Surgeon: Part-II, certification examination case mix. Bone Joint Surg Am 2006, 88:660-667.

[15] Pedowitz RA, Esch J, Snyder S: Evaluation of a virtual reality simulator for arthroscopy skills development. Arthroscopy 2002, 18:E29-E35.

[16] Al-Elq A: Simulation-based medical teaching and learning. J Fam and Com Med 2010, 17:35-40.

[17] Karahan M, Kerkhoffs GM, Randelli P, Tuijthof GJ: Effective Training of Arthroscopic Skills: Springer Berlin Heidelberg, 2015.

[18] Vaughan N, Dubey VN, Wainwright TW, Middleton RG: A review of virtual reality based training simulators for orthopaedic surgery. Med Eng Phys 2015, 38:59-71.

[19] Hui Y, Safir O, Dubrowski A, Carnahan H: What skills should simulation training in arthroscopy teach residents? A focus on resident input. Int J Comput Assist Radiol Surg 2013, 8:945-953.

[20] Safir O, Dubrowski A, Mirksy C, Backstein D, Carnahan H: What skills should simulation training in arthroscopy teach residents? Int J Cars 2008, 3:433-437.

[21] Vitale MA, Kleweno CP, Jacir AM, Levine WN, Bigliani LU, Ahmad CS: Training resources in arthroscopic rotator cuff repair. J Bone Joint Surg Am 2007, 89:1393-1398.

[22] Wolf BR, Britton CL: How orthopaedic residents perceive educational resources. Iowa Orthop J 2013, 33:185-190.

[23] Dunn JC, Belmont PJ, Lanzi J, Martin K, Bader J, Owens B, Waterman BR: Arthroscopic Shoulder Surgical Simulation Training Curriculum: Transfer Reliability and Maintenance of Skill Over Time. J Surg Educ 2015, 72:1118-1123.

[24] Kairys JC, McGuire K, Crawford AG, Yeo CJ: Cumulative operative experience is decreasing during general surgery residency: a worrisome trend for surgical trainees? J Am Coll Surg 2008, 206:804-811.

[25] Hall MP, Kaplan KM, Gorczynski CT, Zuckerman JD, Rosen JE Assessment of arthroscopic training in U.S. orthopedic surgery residency programs--a resident self-assessment. Bull NYU Hosp Jt Dis 2010, 68:5-10.

[26] Cannon WD, Eckhof DG, Garrett Jr. WE, Hunter RE, Sweeney HJ: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res 2006, 442:21-29.

[27] Peabody T: The effect of work hour restrictions on the education of orthopaedic surgery residents. Clin Orthop Relat Res 2006, 449:128-133.

[28] Megali G, Tonet O, Dario P, Vascellari A, Marcacci M: Computer-assisted training system for knee arthroscopy. Int J Medical Robotics and Computer Assisted Surgery 2005, 1:57-66.

[29] Bell RJ: Alternative training models for surgical residency. Surg Clin North Am 2004, 84:13-25.

[30] Kunkler K: The role of medical simulation: an overview. Int J Med Robot 2006, 2:203-210.

[31] Rosen K: The history of medical simulation. J Crit Care 2008, 23:157-166.

[32] Satava R: Accomplishments and challenges of surgical simulation. Dawning of the next-generation surgical education. Surg Endosc 2001, 15:232-241.

[33] Akhtar KS, Chen A, Standfield NJ, Gupte CM: The role of simulation in developing surgical skills. Curr Rev Musculoskelet Med 2014, 7:155-160.

[34] Seropian MA, Brown K, Gavilanes JS, Driggers B: Simulation: Not just a Manikin. J Nurs Educ 2004, 43(4):164-169.

[35] Harewood GC, Murray F, Winder S, Patchett S: Evaluation of formal feedback on endoscopic competence among trainees: the EFFECT trial. Ir J Med Sci 2008, 177:253-256.

[36] O'Connor A, Schwaitzberg SD, Cao CG: How much feedback is necessary for learning to suture? Surg Endosc 2008, 22:1614-1619.

[37] Pinney SJ, Mehta S, Pratt DD, Sarwark JF, Campion E, Blakemore L, Black KP: Orthopaedic surgeons as educators. Applying the principles of adult education to teaching orthopaedic residents. J Bone Joint Surg Am 2007, 89:1385-1392.

[38] Prince M: Does active learning work? A review of the research. J Eng Educ 2004, 93:223-231.

[39] Slade Shantz JA, Leiter JR, Gottschalk T, MacDonald PB: The internal validity of arthroscopic simulators and their effectiveness in arthroscopic education. Knee Surg Sports Traumatol Arthrosc 2014, 22:33-40.

[40] Tuijthof GJ, Horeman T, Schafroth MU, Blankevoort L, Kerkhoffs GMMJ: Probing forces of menisci: what levels are safe for arthroscopic surgery. Knee Surg Sports Traumatol Arthrosc 2011, 19:248-254.

[41] Elliott MJ, Caprise PA, Henning AE, Kurtz CA, Sekiya JK: Diagnostic knee arthroscopy: a pilot study to evaluate surgical skills. Arthroscopy 2012, 28:218-224.

[42] Insel A CB, Leger R, Arciero R, Mazzocca AD: The development of an objective model to assess arthroscopic performance. J Bone Joint Surg Am 2009, 91:2287-2295.

[43] Moorthy K, Smith S, Brown T, Bann S, Darzi A: Evaluation of virtual reality bronchoscopy as a learning and assessment tool. Respiration 2003, 70:195-199.

[44] Brennan BG, Norman GR: Use of encounter cards for evaluation of residents in obstetrics. Acad Med 1997, 72:S43-44.

[45] Hamstra SJ, Dubrowski A: Effective training and assessment of surgical skills, and the correlates of performance. Surg Innov 2005, 12:71-77.

[46] Larson JL, Williams RG, Ketchum J, Boehler ML, Dunnington GL: Feasibility, reliability and validity of an operative performance rating system for evaluating surgery residents. Surgery 2005, 138:640-647.

[47] Norcini JJ, Blank LL, Duffy ED, Fortna GS The mini-CEX: A method for assessing clinical skills Ann Intern Med 2003, 138:476-481.

[48] Rhoton M: A new method to evaluate clinical performance and critical incidents in anaesthesia: Quantification of daily comments by teachers Med Educ 1990, 24:280-289.

[49] Larsen CR, Soerensen JL, Grantcharov TP, Dalsgaard T, Schouenborg L, Ottosen C, Schroeder TV, Ottesen BS: Effect of virtual reality training on laparoscopic surgery: randomised controlled trial. BMJ 2009, 338:B1802-1808.

[50] MacRae H RG, Leadbetter W, Reznick RK: comprehensive examination for senior surgical residents. Am J Surg 2000, 179:190-193.

[51] Martin JA, Rehger G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M: Objective structured assessment of technical skills (OSATS) for surgical residents. Br J Surg 1997, 84:173-178.

[52] Rosen J SM, Hannaford B, Sinanan M: Objective laparoscopic skills assessments of surgical residents using Hidden Markov Models based on haptic information and tool/ tissue interactions. Stud Health Technol Inform 2001, 81:417-423.

[53] Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM: A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005, 190:107-113.

[54] Slade Shantz JA, LJ, Collins JB, MacDonald PB: Validation of a global assessment of arthroscopic skills in a cadaveric knee model. Arthroscopy 2013, 29:106-112.

[55] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner CD, Garrett WE, Olson T, Warme WJ, Nicandri GT: The Arthroscopic Surgical Skill Evaluation Tool (ASSET). Am J Sports Med 2013, 41:1229-1237.

[56] Atesok K, Mabrey JD, Jazrawi LM, Egol KA: Surgical simulation in orthopaedic skills training. J Am Acad Orthop Surg 2012, 20:410-422.

[57] Frank RM, Erickson B, Frank JM, Bush-Joseph CA, Bach BR Jr, Cole BJ, Romeo AA, Provencher MT, Verma NN Utility of modern arthroscopic simulator training models. Arthroscopy 2014, 30:121-133.

[58] Pedowitz RA, Marsch JL: Motor skills training in orthopaedic surgery: a paradigm shift toward a simulation-based educational curriculum. J Am Acad Orthop Surg 2012, 20:407-409.

[59] McCarthy AD, Hollands RJ: A commercially viable virtual reality knee arthroscopy training system. Stud Health Technol Inform 1998, 50:302-308.

[60] Sidhu RS, Park J, Brydges R, MacRae HM, Dubrowski A: Laboratory-based vascular anastomosis training: a randomized controlled trial evaluating the effects of bench model fidelity and level of training on skill acquisition. J Vasc Surg 2007, 45:343-349.

[61] Escoto A, Le Ber F, Trejos AL, Naish MD, Patel RV, Lebel ME: A Knee Arthroscopy simulator: design and validation. Conf Proc IEEE Eng Med Biol Soc 2013: 5715-5718.

[62] Michelson J: Simulation in orthopaedic education: an overview of theory and practice J Bone Joint Surg Am 2006, 88:1405-1411.

[63] Bliss JP, Hanner-Bailey HS, Scerbo MW: Determining the efficacy of an immersive trainer for arthroscopy skills. Stud Health Technol Inform 2005, 111:54-56.

[64] Gurusamy K, Aggarwal R, Palanivelu L and Davidson BR: Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. Br J Surg 2008, 95:1088-1097.

[65] McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR: Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res 2006, 442:13-20.

[66] Sherman KP, Ward JW, Wills DP, Sherman VJ, Mohsen AM: Surgical trainee assessment using a VE knee arthroscopy training system (VE-KATS): experimental results. Stud Health Technol Inform 2001, 81:465-470.

[67] Ström P, Kjellin A, Hedman L, Wredmark T, Felländer-Tsai L:Training in tasks with different visual-spatial components does not improve virtual arthroscopy performance. Surg Endosc 2004, 18:115-120.

[68] Howells NR, Brinsden MD, Gill RS, Carr AJ, Rees JL Motion analysis: a validated method for showing skill levels in arthroscopy. Arthroscopy 2008, 24:335-342.

[69] McCarthy AD, Harley P, Smallwood R: Virtual arthroscopy training: do the virtual skills developed match the real skills required? Stud Health Technol Inform 1999, 62:221-227.

[70] Tashiro Y, Miura H, Nakanishi Y, Okazaki K, Iwamoto Y: Evaluation of skills in arthroscopic training based on trajectory and force data. Clin Orthop Relat Res 2009, 467:546-552.

[71] Tuijthof GJ, van Sterkenburg MN, Sierevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills. Knee Surg Sports Traumatol Arthrosc 2010, 18:218-224.

[72] Obdeijn MC, van Baalen SJ, Horeman T, Liverneaux P, Tuijthof GJM: The Use of Navigation Forces for Assessment of Wrist Arthroscopy Skills Level. J Wrist Surg 2014, 3:132-138.

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

CHAPTER 2

Principles of Validity

The process of validation of training methods and skills assessment tools is important in orthopedic surgery. Based on validity studies, decisions are made about the appropriateness of a method or tool, and whether or not it will be used in a training program. Validity can be defined as ‘‘the property of being true, correct, and in conformity with reality’’[1]. Validity studies clarify whether or not we are we measuring what we want to measure, how well test scores reflect ‘reality’, and if predictions or conclusions predictions based on the measurements are correct. Validation levels have been developed to assess the validity of a test or tool. Increasing in strength of evidence, these include face validity, content validity, construct validity and criterion validity.

Face validity is a qualitative and subjective type of validation and has the lowest strength of evidence. This type of validity is assessed by having experts judge the test to determine the appropriateness of a test and whether the test will measure what it is supposed to measure. It is the first and most basis condition for simulators, concerning the degree of resemblance between the simulator and the actual real life setting[2-5]. Face validity is usually used only during the initial phases of test construction[1].

Content validity is evaluated by examining the contents of the test items; whether the test items are appropriate and overall cohesive, and if a test contains the steps and skills that are used in a procedure. Content validity concerns the degree to which an arthroscopic simulator covers the actual medical psychomotor skills[3]. Like face validity, content validity is mostly subjective, and relies on the judgments of experts about the relevance of the materials used[1].

Construct validity concerns the degree to which the scores of a measurement instrument are consistent with hypotheses. It evaluates whether a test or instrument identifies the quality, ability, or trait it was

designed to measure. Construct validity concerns internal relationships, relationships with scores of other instruments or differences between relevant groups. A simulation task for arthroscopy shows construct validity as an assessment tool if it can distinguish between groups with different levels of arthroscopic experience. *Convergent* and *discriminant* or *divergent validity* are the two subtypes of validity that make up construct validity. Convergent validity refers to the degree to which two measures of constructs that theoretically should be related are in fact related. Discriminant validity tests whether two measures that are supposed to be unrelated are, in fact, unrelated[6]. A form of discriminant validity is *known-groups validity*. Known-groups validity is demonstrated when a test or tool can discriminate between two groups known to differ on the variable of interest[7].

Criterion validity is the degree to which the measurement score is an adequate reflection of a gold standard. Criterion-related validation can be either concurrent or predictive. *Concurrent validity* evaluates to what extent test scores on one instrument and scores on another instrument, that should both measure the same construct, are related[1]. Concurrent validity is demonstrated when high correlation between different assessment tools for arthroscopic skills exist. It can be used when a new assessment tool has to replace an established ‘‘gold standard’’ assessment tool. *Predictive*, or *transfer validity* refers to the extent to which the performance on a test predicts the true (future) outcome in the actual environment[2-5, 8]. A simulator or simulator task has predictive or transfer validity if it is able to predict clinical performance from the simulated performance. A tool used to measure skills will have predictive validity if it predicts who will perform actual surgical tasks well and who will not.

All of these validation levels have value, and are desirable elements of validity in simulation; however, predictive/transfer validity is the strongest evidence of validity and the most likely to provide clinically meaningful information. As perfect validity for any test is not realistic, validation studies aim to explore whether there is enough supporting evidence that teaching and assessment methods teach or test what they are supposed to do[9].

References

[1] Gallagher GE, Ritter M, Satava RM: Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc 2003, 17(10):1525-1529.

[2] Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ: Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005, 19:1523-1532.

[3] Leong JJ LD, Das A, Aggarwal R, Reilly P, Atkinson HD, Emery RJ, Darzi AW: Validation of orthopaedic bench models for trauma surgery. J Bone Joint Surg Br 2008, 90:958-965.

[4] Oropesa I, Sanchez-Gonszalez P, Lamata P, Chmarra MK, Pagador JB, Sánchez-Margallo JA, Sánchez-Margallo FM, Gómez EJ: Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res 2011, 171:E81-E95.

[5] Schijven MP, Jakimowicz J: Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005, 14:244-246.

[6] Campell DT, Fiske DW: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 1959, 56:81-105.

[7] Hattie JA, Cooksey R: Procedures for Assessing the Validities of Tests Using the "Known-Groups" Method. Appl Psy Meas 1984, 8:295-305.

[8] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35:883-888.

[9] Karahan M, Kerkhoffs GM, Randelli P, Tuijthof GJ: Effective Training of Arthroscopic Skills: Springer Berlin Heidelberg, 2015.

CHAPTER 3

How valid are commercially available medical simulators?

J.J. Stunt , P.H.L.M. Wulms , G.M.M.J. Kerkhoffs, J. Dankelman , C.N. van Dijk, G.J.M. Tuijthof

Advances in Medical Education and Practice (2014) 5: 385-395

Abstract

Background: Since simulators offer important advantages, they are increasingly used in medical education and medical skills training that require physical actions. A wide variety of simulators have become commercially available. It is of high importance that evidence is provided that training on these simulators can actually improve clinical performance on live patients. Therefore, the aim of this review is to determine the availability of different types of simulators and the evidence of their validation, to offer insight regarding which simulators are suitable to use in the clinical setting as a training modality.

Summary: Four hundred and thirty-three commercially available simulators were found, from which 405 (94%) were physical models. One hundred and thirty validation studies evaluated 35 (8%) commercially available medical simulators for levels of validity ranging from face to predictive validity. Solely simulators that are used for surgical skills training were validated for the highest validity level (predictive validity). Twenty-four (37%) simulators that give objective feedback had been validated. Studies that tested more powerful levels of validity (concurrent and predictive validity) were methodologically stronger than studies that tested more elementary levels of validity (face, content, and construct validity).

Conclusion: Ninety-three point five percent of the commercially available simulators are not known to be tested for validity. Although the importance of (a high level of) validation depends on the difficulty level of skills training and possible consequences when skills are insufficient, it is advisable for medical professionals, trainees, medical educators, and companies who manufacture medical simulators to critically judge the available medical simulators for proper validation. This way adequate, safe, and affordable medical psychomotor skills training can be achieved.

Keywords: validity level, training modality, medical education, validation studies, medical skills training

This research was funded by the Marti-Keuning Eckhart Foundation, Lunteren, The Netherlands.

Introduction

Simulators for medical training have been used for centuries. More primitive forms of physical models were used before plastic mannequins and virtual systems (VS) were available[1]. Since then, simulation in medical education has been deployed for a variety of actions, such as assessment skills, injections, trauma and cardiac life support, anesthesia, intubation, and surgical skills (SuS)[2, 3]. These actions require psychomotor skills, physical movements that are associated with cognitive processes[4, 5]. Among these psychomotor skills are skills that require (hand–eye) coordination, manipulation, dexterity, grace, strength, and speed. Studies show that medical skills training which requires physical actions can be optimally performed by actual practice in performing these actions, e.g., instrument handling[6]. This is explained by the fact that when learning psychomotor skills, the brain and body co-adapt to improve the manual (instrument) handling. This way, the trainee learns which actions are correct and which are not[5].

Four main reasons to use simulators instead of traditional training in the operating room have been described[6]. Firstly, improved educational experience; when simulators are placed in an easily accessible location, they are available continuously. This overcomes the problem of dependency on the availability of an actual patient case. Simulators also allow easy access to a wide variety of clinical scenarios, e.g., complications[6]. Secondly, patient safety; simulators allow the trainee to make mistakes, which can equip the resident with a basic skills level that would not compromise patient safety when continuing training in the operating room[7-14]. Thirdly, cost efficiency; the costs of setting up a simulation center are in the end often less than the costs of instructors' training time, and resources required as part of the training[6]. Moreover,

the increased efficiency of trainees when performing a procedure adds to the return on investment achieved by medical simulators, as Frost and Sullivan demonstrated[15]. Lastly, simulators offer the opportunity to measure performance and training progress objectively by integrated sensors that can measure, eg, task time, path length, and forces[6, 7, 16-18].

With the increased developments and experiences in research settings, a wide variety of simulators have become commercially available. The pressing question is whether improvements in performance on medical simulators actually translates into improved clinical performance on live patients. Commercially available simulators in other industries, such as aerospace, the military, business management, transportation, and nuclear power, have been demonstrated to be valuable for performance in real life situations[19-23]. Similarly, it is of high importance that medical simulators allow for the correct training of medical skills to improve real life performances. Lack of proper validation could imply that the simulator at hand does not improve skills or worse, could cause incorrect skills training[24, 25].

Since validation of a simulator is required to guarantee proper simulator training, the aim of this review is to determine the availability of medical simulators and whether they are validated or not, and to discuss their appropriateness. This review is distinctive as it categorizes simulators based on simulator type and validation level. In this way, it provides a complete overview of all sorts of available simulators and their degree of validation. This will offer hospitals and medical educators, who are considering the implementation of simulation training in their curriculum, guidelines on the suitability of various simulators to fulfill their needs and demands.

Methods

The approach to achieve the study goal was set as follows. Firstly, an inventory was made of all commercially available simulators that allow medical psychomotor skills training. Secondly, categories that represent medical psychomotor skills were identified and each simulator was placed in one of those categories. Each category will be discussed and illustrated with some representative simulators. Thirdly, validity levels for all available simulators were determined. Lastly, study designs of the validity studies were evaluated in order to determine the reliability of the results of the validity studies.

Inventory of medical simulators

The inventory of commercially available medical simulators was performed by searching the Internet using search engines Google and Yahoo, and the websites of professional associations of medical education (Table 1). The search terms were split up in categories to find relevant synonyms (Table 2). Combinations of these categorized keywords were used as search strategy. For each Internet search engine, a large number of "hits" were found. Relevant websites were selected using the following inclusion criteria: the website needs to be from the company that actually manufactures and sells the product; the simulator should be intended for psychomotor skills training in the medical field (this implies that the models, mannequins or software packages that only offer knowledge training or visualization were excluded); if the company's website provided additional medical simulators, all products that fulfil the criteria were included separately; the website should have had its latest "update" after January 2009, so that it can be expected that the company is still

actively involved in commercial activities in the field of medical simulators.

Table 1 List of societies and associations concerning medical education and simulation. Some of these societies promote commercially available simulators, which were included in our inventory

Abbreviation	Society
SSIH	Society for Simulation in Healthcare
SESAM	Society in Europe of Simulation Applied Medicine
DSSH	Dutch Society for Simulation in Healthcare
INACSL	International Nursing Association for Clinical Nursing Simulation and
NLN	National League for Nursing
ASSH	Australian Society for Simulation in Healthcare
SIRC	Simulation Innovation Resource Center
AMEE	An International Association For Medical Education

Table 2 Search terms

simulator#	medic#	education	Product	Skill#
trainer#	health#	Learning	Company OR firm OR business	psychomotor
"virtual reality" OR VR	Clinical	Teaching	Commerc#	Dexterity
"skills trainer"	Surg#	Training	Purchase OR buy OR offer	Handiness
Model	Nurs#			Eye-hand Coordination
Simulation				
Phantom				
Dummy				
Mannequin				
Manikin				
Mock-up				

Categorization of simulator type

For our study purpose, medical simulators were categorized based on their distinct characteristics: VS and physical mannequins with or without sensors (Figure 1)[14, 26]. VS are software based simulators. The software

simulates the clinical environment that allows practicing individual clinical psychomotor skills. Most of these simulators have a physical interface and provide objective feedback to the user about their performance with task time as the most commonly used performance parameter[26]. The physical mannequins are mostly plastic phantoms simulating (parts of) the human body. The advantage of physical models is that the sense of touch is inherently present, which can provide a very realistic training environment. Most models do not provide integrated sensors and real-time feedback. These models require an experienced professional supervising the skills training. Some physical models have integrated sensors and computer software which allow for an objective performance assessment[14, 27, 28]. As these simulators take over part of the assessment of training progress, it might be expected that they are validated in a different manner. Therefore, a distinction was made between simulators that provide feedback and simulators that do not.

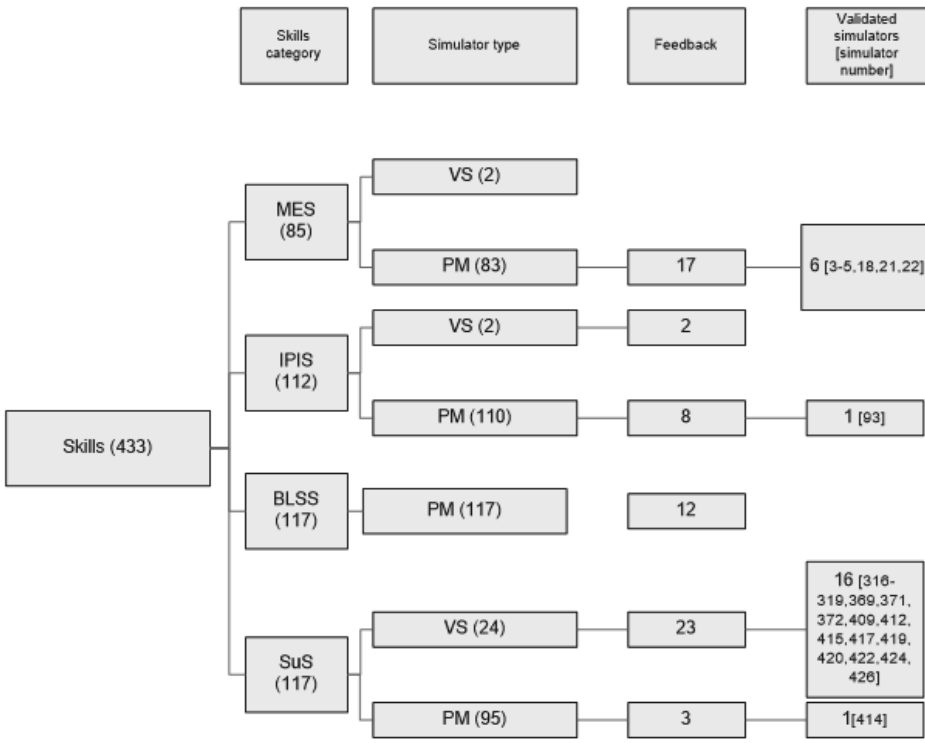


Figure 1: Schematic overview of the number of simulators per skills category (in between brackets) and the number of simulators per simulator type (in between brackets). VS: virtual reality system, and PM: physical model

Categorization of medical psychomotor skills

Skills were categorized in the following categories as they are the most distinct psychomotor skills medical professionals will learn during their education starting at BSc level: 1) manual patient examination skills (MES): an evaluation of the human body and its functions that requires direct physical contact between physician and patient; 2) injections, needle punctures, and intravenous catheterization (peripheral and central) skills (IPIS): the manual process of insertion of a needle into human skin tissue for different purposes such as taking blood samples, lumbar or epidural punctures, injections or vaccinations, or the insertion of a catheter into a vein; 3) basic life support skills (BLSS)[29, 30]. BLSS refers to

maintaining an open airway and supporting breathing and circulation, which can be further divided into the following psychomotor skills: continued circulation, executed by chest compression and cardiac massage; opening the airway, executed by manually tilting the head and lifting the chin; continued breathing, executed by closing the nose, removal of visible obstructions, mouth-to-mouth ventilation, and feeling for breathing[31]; 4) SuS: indirect tissue manipulation for diagnostic or therapeutic treatment by means of medical instruments, eg, scalpels, forceps, clamps, and scissors. Surgical procedures can cause broken skin, contact with mucosa or internal body cavities beyond a natural or artificial body orifice, and are subdivided into minimally invasive and open procedures.

Inventory of validation and study design quality assessment

The brand name of all retrieved simulators added to the keyword "simulator" was used to search PubMed for scientific evidence on validity of that particular simulator. After scanning the abstract, validation studies were included and the level of validation of that particular simulator was noted[32]. Studies were scored for face validity[24, 33] (the most elementary level), construct validity[33], concurrent validity[24, 32], and the most powerful level, predictive validity[24, 32, 33]. The validation studies were evaluated for their study design using Issenberg's guidelines for educational studies involving simulators (Table 3)[34]. Each study was scored for several aspects concerning the research question, participants, methodology, outcome measures, and manner of scoring (Table 4). An outcome measure is considered appropriate when it is clearly defined and measured objectively. The validation studies demonstrated substantial heterogeneity

in study design, therefore, analysis of the data was performed qualitatively and trends were highlighted.

Results

Inventory and categorization of medical simulators

In total, 433 commercially available simulators were found (see Supplementary material: https://c1cnlq18s0hsr7ps1atj.sec.amc.nl/cr_data/supplementary_file_63435.pdf), offered by 24 different companies. From these simulators, 405 (93.5%) are physical models and 28 (6.5%) are virtual simulators (Figure 1). An almost equal distribution of simulators is available for each of the four defined skills categories (Figure 1), with the SuS category containing the noticeably highest portion of virtual reality simulators (86%). Objective feedback was provided by the simulator itself in 65 cases (15%). Simulators for patient examination (MES) training provide the possibility for physical care training, e.g., respiratory gas exchange, intubation, and anesthesia delivery[28, 34-38]. The typical simulators in this category predominantly consist of (full body) mannequins that have anatomical structures and simulate physiological functions such as respiration, and peripheral pulses (eg, Supplementary material: simulators 3 and 21).

IPIS simulators provide training on needle punctures and catheterization. Such simulators usually consist of a mimicked body part, eg, an arm or a torso. An example is the Lumbar Puncture simulator (Kyoto Kagaku Co., Kitanekoyacho Fushimi-ku Kyoto, Japan)[39, 40]. This simulator consists of a life-like lower torso with a removable “skin” that does not show the marks caused by previous needle punctures. Integral to the simulator is a replaceable “puncture block”, which can represent

different types of patients (eg, "normal", "obese", "elderly"), and which is inserted under the "skin"[41].

BLSS simulators allow for emergency care skills training, such as correct head tilt and chin lift, application of cervical collars, splints, and traction or application to spine board[42]. These simulators predominantly consist of full body mannequins having primary features such as anatomically correct landmarks, articulated body parts to manipulate the full range of motion, removable mouthpieces and airways, permitting the performance of chest compressions, oral or nasal intubation, and simulated carotid pulse (eg, Supplementary material: simulators 243, 245, and 270). SuS simulators are used for skills training required when performing open or minimally invasive surgery, like knot tying, suturing, instrument and tissue handling, dissection, simple and complex wound closure. Both physical and virtual simulators form part of this category. A representative example of a physical simulator is the life-sized human torso with thoracic and abdominal cavities and neck/trachea. Such a model is suited to provide training on a whole open surgery procedure, including preparing the operative area, (local) anesthesia, tube insertion, and closure (eg, Supplementary material: 355 and 357). The torso is covered with a polymer that mimics the skin and contains red fluid that mimics blood. Virtual reality systems start to take an important place in minimally invasive surgical procedure training, especially for hand–eye co-ordination training. The VS provide instrument handles with or without a phantom limb and a computer screen in which a virtual scene is presented (eg, the Symbionix simulators 316–322 [Simbionix, Cleveland, OH, USA] and the Simendo simulators 425–426 [Simendo B.V., Rotterdam, the Netherlands] in the Supplementary material). Software

provides a ‘‘plug-and-play’’ connection to a personal computer via a USB port[43, 44].

Inventory of validation and study design quality assessment

One hundred and thirty validation studies evaluated 35 commercially available medical simulators for levels of validity ranging from face to predictive validity (Figure 2). From these 35 simulators, two (5.7%) simulators were tested for face validity, four (11.4%) for content validity, seven (20%) for construct validity, fourteen (40%) for concurrent validity and eight (22.9%) for predictive validity (Figure 2). References of the validated simulators are shown in the Supplementary material (between brackets). Twenty-four (37%) simulators that provide objective feedback have been validated, from which six occurred in MES, one in IPIS and seventeen in SuS (Figure 1).

The numbers of validated simulators per category were substantially different, as was the level of validity (Figure 2). SuS simulators were most validated (62.9%), and most frequently for the highest validity level (Figure 2, predictive validity). MES simulators were primarily tested for content and concurrent validity. The proportion of validated IPIS and BLSS simulators was small (Figure 2).

The quality of the study designs was verified for ten important aspects. Although all studies clearly described the researched question, study population, and outcome measures, few studies met all other criteria on the checklist. Most studies did not perform a power analysis to guarantee a correct number of participants before inclusion. Twelve percent of the 130 studies used a standardized assessment system or performed blind assessment. The majority of the studies (111) performed a correct selection of subjects: either based on experience level or with a

randomly selected control group. However, 20 studies did not select their control group randomly or had no control group at all (Table 3) (37 studies tested face or content validity)[45-48].

Each study used proper outcome measures to test the efficacy of the simulator, which indicated psychomotor skills performance. The most commonly used performance measures are depicted in Table 4. To assess performance data objectively the following standardized scoring methods were used: team leadership-interpersonal skills (TLIS) and emergency clinical care scales (ECCS)[49], objective structural clinical examination (OSCE)[27, 50], objective structured assessment of technical skills (OSAT)[39, 51-55], and global rating scale (GRS)[56-59]. All other studies used assessment methods that were developed specifically for that study.

Methodologically speaking, the studies that tested concurrent and predictive validity outperformed the studies that tested face, content, and construct validity.

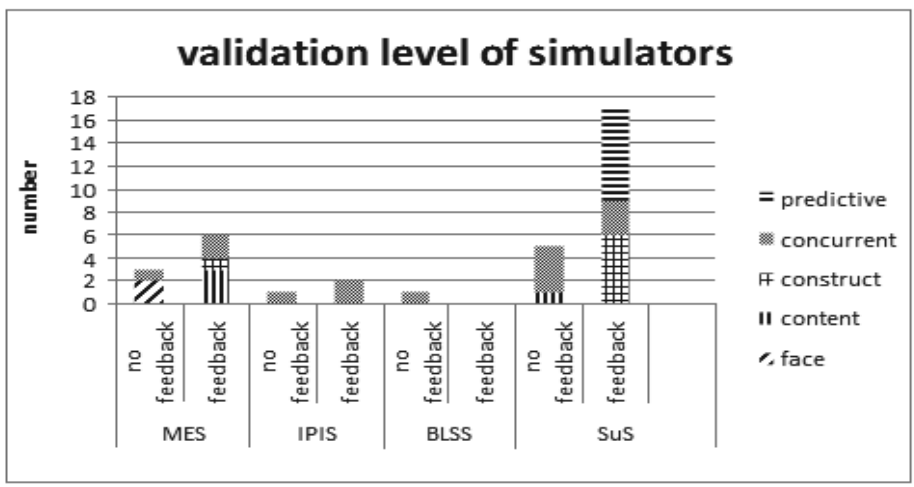


Figure 2: The number of validated simulators. Arrangement is based on skills category, level of validation and whether the simulator give feedback or not. Nine MES simulators, three IPIS simulators, one BLSS simulator and 22 SuS simulators are validated

Discussion

This study reviewed the availability of medical simulators, their validity level, and the reliability of the study designs. Four hundred and thirty-three commercially available simulators were found, of which 405 (94%) were physical models. Evidence of validation was found for 35 (6.5%) simulators (Figure 2). Mainly in category two and three, the number of validated simulators was marginal. Solely SuS simulators were validated for the highest validity level. Sixty-three percent of the 65 simulators that provide feedback on performance have not been validated, which is remarkable as these simulators take over part of the supervisors' judgment. Studies that tested more powerful levels of validity (concurrent and predictive validity) were methodologically stronger than studies that tested more elementary levels of validity (face, content, and construct validity).

These findings can partly be explained: the necessity of a high level validation and the extent to which simulators need to mimic reality is firstly dependent on the type of skills training, and secondly on the possible consequences for patients when medical psychomotor skills are insufficient. This could especially be the case for SuS skills, because minimally invasive SuS are presumably most distinct from daily use of psychomotor skills, and as a result not well developed. In addition, when these skills are taught incorrectly, it can have serious consequences for the patient, e.g., if a large haemorrhage occurs as a result of an incorrect incision. To guarantee patient safety, it is important that simulators designed for this type of training demonstrate high levels of validity[60, 61]. For other types of skills, such as patient examination, a lower validity level can be acceptable, because these skills are closer related to everyday use of psychomotor skills, and solely require a basic level of training on a

simulator, which can be quickly adapted in a real-life situation[45, 46]. Moreover, it requires less extensive methodology to determine face validity than to determine predictive validity.

Certain factors made it difficult to score all validity studies on equal terms; substantial heterogeneity exists among the studies. However, in general, it can be stated that a substantial part of the validation studies showed methodological flaws. For example, many studies did not describe a power analysis, so it was difficult to judge whether these studies included the correct number of participants. Furthermore, only fifteen of 130 studies used standardized assessment methods and blinded assessors. Unvalidated assessment methods and unblinded ratings are less objective, which affects reliability and validity of the test[26]. This raises the question whether the presented studies were adequate enough to determine the validity level of a certain simulator. Future validity studies should focus on a proper study design, in order to increase the reliability of the results.

There are several limitations to our study. Firstly, our inventory of commercially available medical simulators was performed solely by searching the Internet. We did not complement our search by contacting manufacturers or by visiting conferences. This might implicate that our list of available simulators is not complete. Secondly, the available level of validity for the simulators was also determined by searching public scientific databases. Quite possibly, manufacturers have performed validity tests with a small group of experts, but refrained from publishing the results. It is also possible that studies have been rejected for publication or have not been published yet. Therefore, the total number of simulators and number of validated simulators that was found, might be underestimated. However, this does not undermine the fact that few simulators were validated. Especially high levels of validation are scanty.

Our results should firstly make medical trainers aware of the fact that a low number of simulators are actually tested, while validation is truly important. Although it is possible that unvalidated simulators provide proper training, validity of a device is a condition to guarantee proper acquisition of psychomotor skills[1, 6-9, 18] and lack of validity brings the risk of acquisition of improper skills[1, 35]. Secondly, a simulator that provides feedback independent of a professional supervisor, should have been validated to guarantee that the provided feedback is adequate and appropriate in real-life settings[1, 62, 63]. Thirdly, for reliable results of validity studies, proper study design is required. Well conducted studies have shown to be limited so far. Lastly, it is necessary to determine the type of skills educators will offer to their trainees with a simulator and the level of validity that is required to guarantee adequate training.

Our plea is for researchers to collaborate with manufacturers to develop questionnaires and protocols to test newly developed simulators. Simulators from the same category can be tested simultaneously with a large group of relevant participants[43]. When objective evidence for basic levels of validity is obtained, it is important to publish the results so that this information is at the disposal of medical trainers. Before introducing a simulator in the training curriculum, it is recommended to first consider which skills training is needed, and the complexity and possible clinical consequences of executing those skills incorrectly. Subsequently, the minimum required level of validity should be determined for the simulator that allows for that type of skills training. The qualitative results support the concept that the level of validation depends on the difficulty level of skills training and the unforeseen consequences when skills are insufficient or lead to erroneous actions. This combination of selection

criteria should guide medical trainers in the proper selection of a simulator for safe and adequate training.

Table 3: Checklist for the evaluation of validation study, using Issenberg's *Guidelines for educational studies involving simulators*

Guidelines for educational study involving simulators	No of studies	References
1. Clear statement of the research question	130	
2. Clear specification of participants	130	
3. Prospective study	130	
4. Power analysis	10	56, 64-72
5. Random selection of subjects	41	1, 5, 27, 28, 39, 43, 49, 51-57
6. Selection based on experience level	70	14, 43, 44, 50-53, 55, 57, 58, 62, 67, 70-124
7. Is the outcome measure the proper one for the study?	130	
8. Standardized scoring of performance	15	27, 39, 49-54, 56-59, 91, 111, 125
9. Was performance blindly assessed	17	49, 53, 56, 72, 86, 110, 111, 124, 126-134
10. Pre-intervention measurement: yes/no	20	28, 35, 36, 48, 49, 51-53, 63, 66, 82, 83, 102, 103, 135-139

Notes: In the first column, the ten important aspects the studies were evaluated for are stated. The second column shows the number of studies that met the criteria. The third column shows the references of the concerned studies.

Table 4: Outcome measures to test the efficacy of the simulator. The parameters indicate psychomotor skills performance

Parameters	Nr of studies	Reference
Time	53	14, 28, 39, 40, 52, 57, 58, 62, 63, 67, 68, 72, 76-79, 81, 86, 87, 91, 93, 95-103, 105-111, 114, 119, 122-126, 131, 132, 135, 138, 140-144
Path length	15	44, 62, 63, 71, 72, 99, 100, 106, 109, 111, 114, 116, 132, 138
Smoothness	9	71, 72, 111, 112, 115, 116, 132, 134, 144
Number and economy of movement	21	39, 43, 52, 58, 77, 78, 82, 96, 98, 99, 104, 105, 110-112, 115, 125, 127, 138, 140, 142
Number of targets reached or missed	10	2, 73-75, 77, 81, 82, 109, 131, 145
Tissue handling	7	39, 43, 52, 58, 96, 110
Technical skills	9	39, 54, 57, 80, 91, 104, 105, 107, 144
Number of errors or instrument collisions	21	22, 44, 58, 62, 63, 80, 82, 86, 87, 96-99, 104-106, 109, 131, 138, 142, 146
Accuracy	15	71, 72, 96, 97, 102, 104, 111, 112, 115, 116, 123, 124, 126, 128, 132

Conclusion

For correct medical psychomotor skills training and to provide objective and correct feedback it is essential to have a realistic training environment. Scientific testing of simulators is an important way to prove and validate the training method. This review shows that 93.5% of the commercially available simulators are not known to be tested for validity, which implies that no evidence is available that they actually improve individual medical psychomotor skills. From the validity studies that were done for 35 simulators, many show some methodological flaws, which weaken the reliability of the results. It is also advisable for companies that manufacture medical simulators to validate their products and provide scientific evidence to their customers. This way, a quality system becomes available, which contributes to providing adequate, safe, and affordable medical psychomotor skills training.

Disclosure

This research was funded by the Marti-Keuning Eckhart Foundation, Lunteren, the Netherlands. The authors report no conflict of interest.

References

[1] Wilfong DN FD, McKinnon JL, Daniel LH, Wan QC: The effects of virtual intravenous and patient simulator training compared to the traditional approach of teaching nurses: a research project on peripheral i.v. catheter insertion. J Infus Nurs 2011, 34:55-62.

[2] Gorman PJ MA, Krummel TM: Simulation and virtual reality in surgical education: real or unreal? Arch Surg 1999, 134:1203-1208.

[3] Ahmed R NZ, Wolfhagen I: Psychomotor skills for the undergraduate medical curriculum in a developing country--Pakistan. Educ Health 2005, 18:5-13.

[4] Wolpert DM, Ghahramani Z: Computational principles of movement neuroscience. NatNeurosci 2000, 3 Suppl:1212-1217.

[5] Wolpert DM, Ghahramani Z, Flanagan JR: Perspectives and problems in motor learning. Trends Cogn Sci 2001, 5:487-494.

[6] Kunkler K: The role of medical simulation: an overview Int J Med Robot 2006, 2:203-210.

[7] Cannon WD, Eckhoff DG, Garrett Jr. WE, Hunter RE, Sweeney HJ: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res 2006, 442:21-29.

[8] Heng PA, Cheng CY, Wong TT, Wu W, Xu Y, Xie Y, Chui YP, Chan KM, Leung KS: Virtual reality techniques. Application to anatomic visualization and orthopaedics training. Clin Orthop Relat Res 2006, 442:5-12.

[9] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Am 2008, 90:494-499.

[10] McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR: Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res 2006, 442:13-20.

[11] Michelson J: Simulation in orthopaedic education: an overview of theory and practice. J Bone Joint Surg Am 2006, 88:1405-1411.

[12] Poss R, Mabrey JD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Garrett WE Jr, Cannon WD: Development of a virtual reality arthroscopic knee simulator. J Bone Joint Surg Am 2000, 70:192-194.

[13] Safir O, Dubrowski A, Mirksy C, Backstein D, Carnahan H: What skills should simulation training in arthroscopy teach residents? Int J Cars 2008, 3:433-437.

[14] Tuijthof GJ, van Sterkenburg MN, Sierevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills. Knee Surg Sports Traumatol Arthrosc 2010, 18:218-224.

[15] Frost and Sullivan: Return on Investment Study for Medical Simulation Training: Immersion Medical, Inc. Laparoscopy Accutouch System, A Frost & Sullivan Report. http://www.healthleadersmedia.vom/content/138774.pdf. 2012

[16] Insel A CB, Leger R, Arciero R, Mazzocca AD: The development of an objective model to assess arthroscopic performance. J Bone Joint Surg Am 2009, 91:2287-2295.

[17] Tuijthof GJ, Visser P, Sierevelt IN, van Dijk CN, Kerkhoffs GM: Does perception of usefulness of arthroscopic simulators differ with levels of experience? Clin Orthop Relat Res 2011, 469:1701-1708.

[18] Horeman T RS, Jansen FW, Dankelman J, van den Dobbelsteen JJ: Force measurement platform for training and assessment of laparoscopic skills. Surg Endosc 2010, 24:3102-3108.

[19] Issenberg SB MW, Hart IR, Mayer JW, Felner JM, Petrusa ER, Waugh RA, Brown DD, Safford RR, Gessner IH, Gordon DL, Ewy GA: Simulation technology for health care professional skills training and assessment. JAMA 1999, 282:861-866.

[20] Goodman W: The world of civil simulators. Flight International Magazine 1978, 18:435.

[21] Streufert S PR, Piasecki M: Simulation-based assessment of managerial competence: reliability and validity. Personnel Psychology 1988, 41:537-557.

[22] Rolfe VJM SK: Flight simulation, 1988.

[23] Keys B WJ: The Role of Management Games and Simulations in Education and Research. Journal of Management 1990, 16:307-336.

[24] Schijven MP, Jakimowicz JJ: Validation of virtual reality simulators: Key to the successful integration of a novel teaching technology into minimal access surgery. Minim Invasive Ther Allied Technol 2005, 14:244-246.

[25] Chmarra MK, Klein S, de Winter JC, Jansen FW, Dankelman J: Objective classification of residents based on their psychomotor laparoscopic skills. Surg Endosc 2010, 24:1031-1039.

[26] van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J: Objective assessment of technical surgical skills. Br J Surg 2010, 97:972-987.

[27] Alinier G, Hunt WB, Gordon R: Determining the value of simulation in nurse education: study design and initial results. Nurse Educ Pract 2004, 4:200-207.

[28] Donoghue AJ, Durbin DR, Nadel FM, Stryjewski GR, Kost SI, Nadkarni VM: Effect of high-fidelity simulation on Pediatric Advanced Life Support training in pediatric house staff: a randomized trial. Pediatr Emerg Care 2009, 25:139-144.

[29] Monsieurs KG, De Regge M, Schelfout S, D'Hondt F, Mpotos N, Valcke M, Calle PA: Efficacy of a self-learning station for basic life support refresher training in a hospital: a randomized controlled trial. Eur J Emerg Med 2012, 19:214-219.

[30] European Society for Emergency Medicine. Emergency Medicine http://www.eusem.org/whatisem/. 2012.

[31] Handley A: Basic life support. Br J Anaesth 1997, 79:151-158.

[32] Oropesa I, Sanchez-Gonzalez P, Lamata P, Chmarra MK, Pagador JB, Sánchez-Margallo JA, Sánchez-Margallo FM, Gómez EJ Methods and tools for objective assessment of psychomotor skills in laparoscopic surgery. J Surg Res 2011, 171:E81-E95.

[33] Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ: Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005, 19:1523-1532.

[34] Issenberg SB, McGaghie WC, Waugh RA: Computers and evaluation of clinical competence. Ann Intern Med 1999, 130:244-245.

[35] Martin JT, Reda H, Dority JS, Zwischenberger JB, Hassan ZU: Surgical resident training using real-time simulation of cardiopulmonary bypass physiology with echocardiography. J Surg Educ 2011, 68:542-546.

[36] Tan GM, Ti LK, Suresh S, Ho BS, Lee TL Teaching first-year medical students physiology: does the human patient simulator allow for more effective teaching? Singapore Med J 2002, 43:238-242.

[37] Nehring WM, Lashley FR: Current use and opinions regarding human patient simulators in nursing education: an international survey. Nurs Educ Perspect 2004 25:244-248.

[38] Fernandez GL, Lee PC, Page DW, D'Amour EM, Wait RB, Seymour NE: Implementation of full patient simulation training in surgical residency. J Surg Educ 2010, 67:393-399.

[39] Bath J, Lawrence P, Chandra A, O'Connell J, Uijtdehaage S, Jimenez JC, Davis G, Hiatt J: Standardization is superior to traditional methods of teaching open vascular simulation. J Vasc Surg 2011, 53:229-234.

[40] Uppal V, Kearns RJ, McGrady EM: Evaluation of M43B Lumbar puncture simulator-II as a training tool for identification of the epidural space and lumbar puncture. Anaesthesia 2011, 66:493-496.

[41] Uppal JK VR, Hazari PP, Chuttani K, Kaushik NK, Mishra AK: Biological evaluation of avidin-based tumor pretargeting with DOTA-Triazole-Biotin constructed via versatile Cu(I) catalyzed click chemistry. J Drug Target 2011, 19:418-426.

[42] Laerdal, helping saves lives. http://www.laerdal.com/. 2012.

[43] Verdaasdonk EG SL, Monteny LJ, Dankelman J: Standardization is superior to traditional methods of teaching open vascular simulation. Surg Endosc 2006, 20:511-518.

[44] Verdaasdonk EG, Stassen LP, Schijven MP, Dankelman J: Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. SurgEndosc 2007, 21:1406-1412.

[45] Verma A, Bhatt H, Booton P, Kneebone R: The Ventriloscope(R) as an innovative tool for assessing clinical examination skills: appraisal of a novel method of simulating auscultatory findings. Med Teach 2011, 33:e388-e396.

[46] Wilson M, Shepherd I, Kelly C, Pitzner J: Assessment of a low-fidelity human patient simulator for the acquisition of nursing skills. Nurse Educ Today 2005, 25:56-57.

[47] Kruger A, Gillmann B, Hardt C, Doring R, Beckers SK, Rossaint R Teaching non-technical skills for critical incidents: Crisis resource management training for medical students. Anaesthesist 2009, 58:582-588.

[48] Shukla A, Kline D, Cherian A, Lescanec A, Rochman A, Plautz C, Kirk M, Littlewood KE, Custalow C, Srinivasan R, Martin ML A: simulation course on lifesaving techniques for third-year medical students. Simul Healthc 2007, 2:11-15.

[49] Pascual JL, Holena DN, Vella MA, Palmieri J, Sicoutris C, Selvan B, Fox AD, Sarani B, Sims C, Williams NN, Schwab CW: Short simulation training improves objective skills in established advanced practitioners managing emergencies on the ward and surgical intensive care unit. J Trauma 2011, 71:330-337.

[50] Karnath B, Thornton W, Frye AW: Teaching and testing physical examination skills without the use of patients. Acad Med 2002 77:753-755.

[51] Nadler I, Liley HG, Sanderson PM: Clinicians can accurately assign Apgar scores to video recordings of simulated neonatal resuscitations. Simul Healthc 2010, 5:204-212.

[52] Chou DS, Abdelshedid C, Clayman RV, McDougall EM: Comparison of results of virtual-reality simulator and training model for basic ureteroscopy training. J Endourol 2006, 20:266-271.

[53] Lucas S, Tuncel A, Bensalah K, Zeltser I, Jenkins A, Pearle M, Cadeddu J: Virtual reality training improves simulated laparoscopic surgery performance in laparoscopy naive medical students. J Endourol 2008, 22:1047-1051.

[54] Korets R, Mues AC, Graversen JA, Gupta M, Benson MC, Cooper KL, Landman J, Badani KK: Validating the use of the Mimic dV-trainer for robotic surgery skill acquisition among urology residents. Urology 2011, 78:1326-1330.

[55] Adler MD, Trainor JL, Siddall VJ, McGaghie WC: Development and evaluation of high fidelity simulation case scenarios for pediatric resident education. Ambul Pediatr 2007, 7:182-186.

[56] Schout BM, Ananias HJ, Bemelmans BL, d'Ancona FC, Muijtjens AM, Dolmans VE, Scherpbier AJ, Hendrikx AJ: Transfer of cysto-urethroscopy skills from a virtual-reality simulator to the operating room: a randomized controlled trial. BJU Int 2010, 106:226-231.

[57] Schout BM, Muijtjens AM, Hendrikx AJ, Ananias HJ, Dolmans VE, Scherpbier AJ, Bemelmans BL: Acquisition of flexible cystoscopy skills on a virtual reality simulator by experts and novices. BJU Int 2010, 105:234-239.

[58] Knoll T, Trojan L, Haecker A, Alken P, Michel MS: Validation of computer-based training in ureterorenoscopy. BJU Int 2005, 95:1276-79.

[59] Wilhelm DM, Ogan K, Roehrborn CG, Cadeddu JA, Pearle MS. Assessment of basic endoscopic performance using a virtual reality simulator. J Am Coll Surg. 2002 ,195(5):675-81

[60] Inspectie voor de Gezondheidszorg. Rapport 'Risico's minimaal invasieve chirurgie onderschat, kwaliteitssysteem voor laparoscopische operaties ontbreekt'. Healthcare Inspectorate. Risks minimally invasive surgery underestimated. http://www.igz.nl/zoeken/document.aspx?doc=Rapport+'Risic 2012.

[61] Issenberg SB, McGaghie WC, Petrusa ER, Lee Gordon D, Scalese RJ: Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005, 27(1):10-28.

[62] Rosenthal R, Gantert WA, Hamel C, Hahnloser D, Metzger J, Kocher T, Vogelbach P, Scheidegger D, Oertli D, Clavien PA: Assessment of construct validity of a virtual reality laparoscopy simulator. J Laparoendosc Adv Surg Tech A 2007, 17:407-513.

[63] Kuduvalli PM JA, Tighe SQ, Robin NM: Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Anaesthesia 2008, 63:364-369.

[64] Oddone EZ, Waugh RA, Samsa G, Corey R, Feussner JR: Teaching cardiovascular examination skills: results from a randomized controlled trial. Am J Med 1993, 95:389-396.

[65] Jones JS, Hunt SJ, Carlson SA, Seamon JP: Assessing bedside cardiologic examination skills using "Harvey," a cardiology patient simulator. Acad Emerg Med 1997, 4:980-985.

[66] Ende A, Zopf Y, Konturek P, Naegel A, Hahn EG, Matthes K, Maiss J: Strategies for training in diagnostic upper endoscopy: a prospective, randomized trial. Gastrointest Endosc 2012, 75:254-260.

[67] Koch AD, Buzink SN, Heemskerk J, Botden SM, Veenendaal R, Jakimowicz JJ, Schoon EJ: Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 2008, 22:158-162.

[68] Buzink SN, Goossens RH, Schoon EJ, de RH, Jakimowicz JJ: Do basic psychomotor skills transfer between different image-based procedures? World J Surg 2010, 34:933-940.

[69] Kothari SN, Kaplan BJ, DeMaria EJ, Broderick TJ, Merrell RC: Training in laparoscopic suturing skills using a new computer-based virtual reality simulator (MIST-VR) provides results comparable to those with an established pelvic trainer system. J Laparoendosc Adv Surg Tech A 2002, 12:167-173.

[70] Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. J Shoulder Elbow Surg 2004, 13:196-205.

[71] Chandra V, Nehra D, Parent R, Woo R, Reyes R, Hernandez-Boussard T, Dutta S: A comparison of laparoscopic and robotic assisted suturing performance by experts and novices. Surgery 2010, 147:830-839.

[72] Neary PC, Boyle E, Delaney CP, Senagore AJ, Keane FB, Gallagher AG: Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. Surg Endosc 2008, 22:2301-2309.

[73] Ferlitsch A, Glauninger P, Gupper A, Schillinger M, Haefner M, Gangl A, Schoefl R: Evaluation of a virtual endoscopy simulator for training in gastrointestinal endoscopy. Endoscopy 2002, 34:698-702.

[74] Ritter EM, McClusky DA, III, Lederman AB, Gallagher AG, Smith CD: Objective psychomotor skills assessment of experienced and novice flexible endoscopists with a virtual reality simulator. J Gastrointest Surg 2003, 7:871-877.

[75] Eversbusch A, Grantcharov TP: Learning curves and impact of psychomotor training on performance in simulated colonoscopy: a randomized trial using a virtual reality endoscopy trainer. Surg Endosc 2004, 18:1514-1518.

[76] Felsher JJ, Olesevich M, Farres H, Rosen M, Fanning A, Dunkin BJ, Marks JM: Validation of a flexible endoscopy simulator. Am J Surg 2005, 189:497-500.

[77] Grantcharov TP, Carstensen L, Schulze S: Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. JSLS 2005, 9:130-133.

[78] Fayez R, Feldman LS, Kaneva P, Fried GM: Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters. Surg Endosc 2010, 24:1060-1065.

[79] Gettman MT, Le CQ, Rangel LJ, Slezak JM, Bergstralh EJ, Krambeck AE: Analysis of a computer based simulator as an educational tool for cystoscopy: subjective and objective results. J Urol 2008, 179:267-271.

[80] Ogan K JL, Shulman MJ, Roehrborn CG, Cadeddu J, Pearle MS: Virtual Ureteroscopy Predicts Ureteroscopic Proficiency of Medical Students on a Cadaver. J Urol 2004, 172:667-671.

[81] Shah J, Darzi A: Virtual reality flexible cystoscopy: a validation study BJU Int 2002, 90:828-832.

[82] Mishra S, Kurien A, Patel R, Patil P, Ganpule A, Muthu V, Sabnis RB, Desai M: Validation of virtual reality simulation for percutaneous renal access training. J Endourol 2010, 24:635-640.

[83] Zhang A, Hunerbein M, Dai Y, Schlag PM, Beller S: Construct validity testing of a laparoscopic surgery simulator (Lap Mentor): evaluation of surgical skill with a virtual laparoscopic training simulator. Surg Endosc 2008, 22:1440-1444.

[84] Yamaguchi S, Konishi K, Yasunaga T, Yoshida D, Kinjo N, Kobayashi K, Ieiri S, Okazaki K, Nakashima H, Tanoue K, Maehara Y, Hashizume M: Construct validity for eye-hand coordination skill on a virtual reality laparoscopic surgical simulator. Surg Endosc 2007, 21:2253-2257.

[85] Ayodeji ID, Sschijven M, Jakimowicz J, Greve JW: Face validation of the Simbionix LAP Mentor virtual reality training module and its applicability in the surgical curriculum. Surg Endosc 2007, 21:1641-1649.

[86] Dayan AB, Ziv A, Berkenstadt H, Munz Y: A simple, low-cost platform for basic laparoscopic skills training. Surg Innov 2008, 15:136-142.

[87] Boon JR, Salas N, Avila D, Boone TB, Lipshultz LI, Link RE: Construct validity of the pig intestine model in the simulation of laparoscopic urethrovesical anastomosis: tools for objective evaluation. J Endourol 2008, 22:2713-2716.

[88] Taffinder N, Sutton C, Fishwick RJ, McManus IC, Darzi A: Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: results from randomised controlled studies using the MIST VR laparoscopic simulator. Stud Health Technol Inform 1998, 50:124-130.

[89] McNatt SS, Smith CD: A computer-based laparoscopic skills assessment device differentiates experienced from novice laparoscopic surgeons. Surg Endosc 2001, 15:1085-1089.

[90] Maithel S, Sierra R, Korndorffer J, Neumann P, Dawson S, Callery M, Jones D, Scott D: Construct and face validity of MIST-VR, Endotower, and CELTS: are we ready for skills assessment using simulators? Surg Endosc 2006, 20:104–12.

[91] Botden SM, de Hingh IH, Jakimowicz JJ: Meaningful assessment method for laparoscopic suturing training in augmented reality. Surg Endosc 2009, 23:2221-2228.

[92] Brewin J NT, Challacombe B, Elhage O, Keisu J, Dasgupta P: Face, content and construct validation of the first virtual reality laparoscopic nephrectomy simulator. BJU Int 2010, 106:850-854.

[93] Schijven MP, Jakimowicz JJ: The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc 2004, 18:121-127.

[94] Schijven MP, Jakimowicz JJ: Face-, expert, and referent validity of the Xitact LS500 laparoscopy simulator. Surg Endosc 2002, 16:1764-1770.

[95] Schijven MP, Jakimowicz JJ: Construct validity: experts and novices performing on the Xitact LS500 laparoscopy simulator. Surg Endosc 2003, 17:803-810.

[96] Bajka M, Tuchschmid S, Streich M, Fink D, Szekely G, Harders M: Evaluation of a new virtual-reality training simulator for hysteroscopy. Surg Endosc 2009, 23:2026-2033.

[97] Bajka M, Tuchschmid S, Fink D, Szekely G, Harders M: Establishing construct validity of a virtual-reality training simulator for hysteroscopy via a multimetric scoring system. Surg Endosc 2010, 24:79-88.

[98] Buzink SN, Botden SM, Heemskerk J, Goossens RH, de RH, Jakimowicz JJ: Camera navigation and tissue manipulation; are these laparoscopic skills related? Surg Endosc 2009, 23:750-757.

[99] Buzink SN, Goossens RH, de Ridder H, Jakimowicz JJ: Training of basic laparoscopy skills on SimSurgery SEP. Minim Invasive Ther Allied Technol 2010, 19:35-41.

[100] van der Meijden OA, Broeders IA, Schijven MP: The SEP "robot": a valid virtual reality robotic simulator for the Da Vinci Surgical System? Surg Technol Int 2010, 19:51-58.

[101] Wohaibi EM, Bush RW, Earle DB, Seymour NE: Surgical resident performance on a virtual reality simulator correlates with operating room performance. J Surg Res 2010, 160:67-72.

[102] Fichera A, Prachand V, Kives S, Levine R, Hasson H Physical reality simulation for training of laparoscopists in the 21st century. A multispecialty, multi-institutional study. JSLS 2005, 9:125-9.

[103] Hasson HM, Kumari NV, Eekhout J: Training simulator for developing laparoscopic skills. JSLS 2001, 5:255-256.

[104] Hung AJ, Zehnder P, Patil MB, Cai J, Ng CK, Aron M, Gill IS, Desai MM: Face, content and construct validity of a novel robotic surgery simulator. J Urol 2011, 186:1019-1024.

[105] Lerner MA, Ayalew M, Peine WJ, Sundaram CP: Does training on a virtual reality robotic simulator improve performance on the da Vinci surgical system? J Endourol 2010, 24:467-472.

[106] Kenney PA, Wszolek MF, Gould JJ, Libertino JA, Moinzadeh A: Face, content, and construct validity of dV-trainer, a novel virtual reality simulator for robotic surgery. Urology 2009, 73:1288-1292.

[107] Iwata N, Fujiwara M, Kodera Y, Tanaka C, Ohashi N, Nakayama G, Koike M, Nakao A: Construct validity of the LapVR virtual-reality surgical simulator. Surg Endosc 2011, 25:423-428.

[108] Sethi AS, Peine WJ, Mohammadi Y, Sundaram CP: Validation of a novel virtual reality robotic simulator J Endourol 2009, 23:503-508.

[109] Mahmood T, Darzi A: A study to validate the colonoscopy simulator. Surg Endosc 2003, 17:1583–1589.

[110] Broe D, Ridgway PF, Johnson S, Tierney S, Conlon KC Construct validation of a novel hybrid surgical simulator. Surg Endosc 2006, 20:900-904.

[111] Van Sickle KR, McClusky DA 3rd, Gallagher AG, Smith CD: Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 2005, 19:1227-1231.

[112] Pellen MG, Horgan LF, Barton JR, Attwood SE: Construct validity of the ProMIS laparoscopic simulator. Surg Endosc 2009, 23:130-139.

[113] Botden SM, Berlage JT, Schijven MP, Jakimowicz JJ: Face validity study of the ProMIS augmented reality laparoscopic suturing simulator Surg Technol Int 2008, 17:26-32.

[114] Jonsson MN, Mahmood M, Askerud T, Hellborg H, Ramel S, Wiklund NP, Kjellman M, Ahlberg G: ProMIS can serve as a da Vinci(R) simulator--a construct validity study. J Endourol 2011, 25:345-350.

[115] Feifer A, Delisle J, Anidjar M: Hybrid augmented reality simulator: preliminary construct validation of laparoscopic smoothness in a urology residency program. J Urol 2008, 180:1455-1459.

[116] Ritter EM, Kindelan TW, Michael C, Pimentel EA, Bowyer MW: Concurrent validity of augmented reality metrics applied to the fundamentals of laparoscopic surgery (FLS). Surg Endosc 2007, 21:1441-1445.

[117] Woodrum DT, Adreatta PB, Yellamanchilli RK, Feryus L, Gauger PG, Minter RM: Construct validity of the LapSim laparoscopic surgical simulator. Am J Surg 2006, 161:28-32.

[118] Duffy AJ, Hogle NJ, McCarthy H, Lew JI, Egan A, Christos P, Fowler DL: Construct validity for the LAPSIM laparoscopic surgical simulator. Surg Endosc 2005, 19:401-405.

[119] Dayal R, Faries PL, Lin SC, Bernheim J, Hollenbeck S, DeRubertis B, Trocciola S, Rhee J, McKinsey J, Morrissey NJ, Kent KC: Computer simulation as a component of catheter-based training. J Vasc Surg 2004, 40:1112-1117.

[120] Nicholson WJ, Caes CU, Patel AD, Niazi K, Palmer S, Helmy T, Gallagher AG: Face and content validation of virtual reality simulation for carotid angiography: results from the first 100 physicians attending the Emory NeuroAnatomy Carotid Training (ENACT) program. Simul Healthc 2006, 1:147-150.

[121] Willoteaux S LC, Duhamel A, Vernhet H, Sapoval M, Boyer L, Bartoli JM, Rousseau H, Beregi JP: Virtual interventional radiology: evaluation of performances as a function of experience. J Radiol 2009, 90:37-41.

[122] Van Herzeele I, Aggarwal R, Neequaye S, Darzi A, Vermassen F, Cheshire NJ: Cognitive training improves clinically relevant outcomes during simulated endovascular procedures. J Vasc Surg 2008, 48:1223-1230.

[123] Berry M, Reznick R, Lystig T, Lonn L: The use of virtual reality for training in carotid artery stenting: a construct validation study. Acta Radiol 2008, 49:801-305.

[124] Tedesco MM, Pak JJ, Harris EJ Jr., Krummel TM, Dalman RL, Lee JT: Simulation-based endovascular skills assessment: the future of credentialing? J Vasc Surg 2008, 47:1008-111.

[125] Van Sickle KR, Buck L, Willis R, Mangram A, Truitt MS, Shabahang M, Thomas S, Trombetta L, Dunkin B, Scott D: A multicenter, simulation-based skills training collaborative using shared GI Mentor II systems: results from the Texas Association of Surgical Skills Laboratories (TASSL) flexible endoscopy curriculum. Surg Endosc 2011, 25:2980-2986.

[126] Shirai Y, Yoshida T, Shiraishi R, Okamoto T, Nakamura H, Harada T, Nishikawa J, Sakaida I: Prospective randomized study on the use of a computer-based endoscopic simulator for training in esophagogastroduodenoscopy. J Gastroenterol Hepatol 2008, 23:1046-1050.

[127] Andreatta PB, Woodrum DT, Birkmeyer JD, Yellamanchilli RK, Doherty GM, Gauger PG, Minter RM: Laparoscopic skills are improved with LapMentor training: results of a randomized, double-blinded study. Ann Surg 2006, 243:854-860.

[128] Youngblood PL, Srivastava S, Curet M, Heinrichs WL, Dev P, Wren SM: Comparison of training on two laparoscopic simulators and assessment of skills transfer to surgical performance. J Am Coll Surg 2005, 200:546-551.

[129] Seymour N, Gallagher AG, Roman, SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM: Virtual Reality Training Improves Operating Room Performance Results of a Randomized, Double-Blinded Study. Annals of surgery 2002, 236:458-464.

[130] Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004, 91:146-150.

[131] Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D: Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? Endoscopy 2005, 17:1583-1589.

[132] Cesanek P, Uchal M, Uranues S, Patruno J, Gogal C, Kimmel S, Bergamaschi R: Do hybrid simulator-generated metrics correlate with content-valid outcome measures? Surg Endosc 2008, 22:2178-2183.

[133] Verdaasdonk EG, Dankelman J, Lange JF, Stassen LP: Transfer validity of laparoscopic knot-tying training on a VR simulator to a realistic environment: a randomized controlled trial. Surg Endosc 2008, 22:1636-1642.

[134] Sotto JA, Ayuste EC, Bowyer MW, Almonte JR, Dofitas RB, Lapitan MC, Pimentel EA, Ritter EM, Wherry DC: Exporting simulation technology to the Philippines: a comparative study of traditional versus simulation methods for teaching intravenous cannulation. Stud Health Technol Inform 2009, 142:346-351.

[135] Shah J, Montgomery B, Langley S, Darzi A: Validation of a flexible cystoscopy course. BJU Int 2002, 90:833-885.

[136] Hall A: Randomized objective comparison of live tissue training versus simulators for emergency procedures. Am Surg 2011, Am Surg 5.

[137] Debes AJ, Aggarwal R, Balasundaram I, Jacobsen MB A tale of two trainers: virtual reality versus a video trainer for acquisition of basic laparoscopic skills. Am J Surg 2010, 199:840-845.

[138] Torkington J, Smith SG, Rees B, Darzi A: The role of the basic surgical skills course in the acquisition and retention of laparoscopic skill. Surg Endosc 2001, 15:1071-1075.

[139] St Clair EW, Oddone EZ, Waugh RA, Corey GR, Feussner JR: Assessing housestaff diagnostic skills using a cardiology patient simulator. Ann Intern Med 1992, 117:751-756.

[140] Vitish-Sharma P, Knowles J, Patel B: Acquisition of fundamental laparoscopic skills: is a box really as good as a virtual reality trainer? Int J Surg 2011, 9:659-661.

[141] Bayona S, Fernández-Arroyo JM, Martin I, Bayona P: Assessment study of insightARTHRO VR arthroscopy virtual training simulator: face, content, and construct validities. J Robotic Surg 2008, 2:151-158.

[142] Rowe R, Cohen RA: An evaluation of a virtual reality airway simulator. Anesth Analg 2002, 95 (1): 62-66.

[143] Van Herzeele I, Aggarwal R, Choong A, Brightwell R, Vermassen FE, Cheshire NJ: Virtual reality simulation objectively differentiates level of carotid stent experience in experienced interventionalists. J Vasc Surg 2007, 46:855-863.

[144] Patel AD, Gallagher AG, Nicholson WJ, Cates CU: Learning curves and reliability measures for virtual reality simulation in the performance assessment of carotid angiography. J Am Coll Cardiol 2006, 47:1796-1802.

[145] Cavaleiro AP, Guimaraes H, Calheiros F: Training neonatal skills with simulators? Acta Paediatr 2009, 98:636-639.

[146] Giovanni de GD, Roberts T, Norman G: Relative effectiveness of high- versus low-fidelity simulation in learning heart sounds. Med Educ 2009, 43:661-668.

CHAPTER 4

Validation of the PASSPORT V2 training environment for arthroscopic skills

J.J. Stunt , G.M.M.J. Kerkhoffs , T. Horeman, C.N. van Dijk, G.J.M. Tuijthof

Knee Surg Sports Traumatol Arthrosc (2014) 24(6):2038-45

Abstract

Purpose: Virtual reality simulators used in the education of orthopaedic residents have often no realistic force feedback. Therefore, a new design, the PASSPORT, was developed. Recently, it has been subject to fundamental changes. The purpose of this study is to demonstrate its face and construct validity, using a previously used protocol.

Methods & Materials: Thirty-one participants were divided into three groups having different levels of arthroscopic experience. Participants answered questions regarding general information and the outer appearance of the simulator for face validity. Construct validity was assessed with one standardized navigation task. Face validity, educational value and user friendliness were further determined with two exercises and by asking participants to fill out the questionnaire. A value of 7 or greater was considered sufficient.

Results: Construct validity was demonstrated between experts and novices. Median task times of the first trial were 137 (57-240) seconds for novices, 98.5 (37-166) seconds for intermediates, and 69 (52-109) seconds for experts. Median task times were not significantly different for most repetitions between the novices and intermediates, and for all repetitions between intermediates and experts. Face validity, educational value and user-friendliness were perceived as sufficient (median > 7). The presence of realistic tactile feedback was considered the biggest asset of the simulator.

Conclusion: The PASSPORT showed construct- and face validity. The PASSPORT is a suitable preparation for real-life arthroscopy, although there is room for improvement. Proper preparation for arthroscopic operations will increase the quality of real life surgery and patients safety.

This research was funded by the Marti-Keuning Eckhart Foundation, Lunteren, The Netherlands.

Introduction

As arthroscopic skills are complex, acquiring them is challenging[1], while training time is limited[2, 3] and demands for patient safety and quality control are increasing. Therefore, simulators were introduced in the education of orthopedic residents. Initially, low-fidelity simulators were used. These type of simulators provide good basic skill development, but as they do not offer realistic simulation, they are not optimal to enhance training[4]. Virtual reality simulators were introduced to optimize orthopedic education[3, 5-8]. Virtual reality simulators have a more developed technology and are more realistic than low fidelity simulators. Previous studies of virtual simulators have demonstrated certain types of validity[3, 5, 9-27]. However, a serious drawback of virtual simulators is their lack of realistic force feedback[10, 28]. Realistic force feedback is essential to imitate an arthroscopic procedure[29-31] and when absent it reduces the natural sense of interaction of the instrument with the tissue[32]. This might be a contributing factor for the fact that training on (virtual reality) simulators for arthroscopic skills has not been adopted as in other endoscopic disciplines.

Within our group, a different approach was developed for the design of an arthroscopic knee simulator that potentially does offer complete natural sensory feedback, while maintaining the advantage of virtual reality simulators. In 2009, the first prototype of this physical simulation environment to Practice Arthroscopic Surgical Skills for Perfect Operative Real-life Treatment (PASSPORT) was presented, and its potential to evolve as a training modality for knee arthroscopy was demonstrated[33]. Although both construct and face validity were demonstrated, substantial improvements were required[33]. Face validity can be defined as the ability of a test to resemble the real situation. Construct validity is the ability of a

test to actually assess what it is designed for[34]. Over the past years, the PASSPORT has been subjected to the necessary major changes in appearance, registration of performance and interaction with the trainee (Fig. 1). These changes are fundamental and are expected to contribute significantly to the PASSPORT concept, especially in terms of face validity and the potential to practice arthroscopic skills beyond navigation and probing.

The purpose of this study is to demonstrate face and construct validity of PASSPORT V2 and to show that this simulator has enhanced realism compared to virtual reality simulators. This is an important asset as realistic force feedback is essential in arthroscopic training. Moreover, findings of this study will provide clinical relevant information for hospitals that wish to use a simulator for their training program. The study will be performed by using a testing protocol that has been previously applied for validation of other arthroscopic simulators[35].

Materials and Methods

Prototype of PASSPORT V2

The overall concept of PASSPORT remained: The surgical setting is imitated as close as possible by using standard arthroscopic equipment; the human joint is replaced with a phantom model, and sensors are integrated to provide feedback and registration of training sessions[33]. However, the PASSPORT V2 is substantially redesigned compared to the first prototype PASSPORT V1[33, 36]. First the appearance was substantially improved. The original metal rod representing the lower leg was replaced by a dummy leg of a mannequin (Fig. 1.D). A patellar bone was added that moves in line with the flexion-extension motion of the lower leg (Fig. 1.G). The tibia plateau (Fig. 1.M), the menisci (Fig. 1.L) and cruciate

ligaments were upgraded. The menisci were attached such that they could be replaced by two simple actions (Fig. 1.F). A special hinge was designed (Fig. 1.J), which allows both flexion-extension and joint stressing in a natural manner[36]. Springs provide the resistance against stressing. Their stiffness was chosen such that in the hinge's end position, the required stressing force was around 78 N[37].

Second, to measure the performance of a trainee, a set of sensors was incorporated in the system in such a manner that these sensors do not influence the intra-articular joint face validity. Motion is measured in a two-dimensional plane using a webcam and yellow-colored markers that are attached to the arthroscope and the instruments (Fig. 1.A and C). The choice to measure motion in a plane was made, as it is assumed that due to the confined knee joint space motion predominantly takes place in one plane, and because task time is strongly correlated with path length[38-40]. Additionally, one 3D force sensor was connected to the tibia plateau and one 3D force sensor to the femur to measure instrument-tissue interaction on the delicate cartilage and menisci areas (Fig. 1.N). The sensors were based on our force measurement platform[41].

Third, a graphical user interface was added that leads trainees through a range of exercises, indicates progress in training by means of spider graphs and gives real-time feedback on performance during exercises (Fig. 1.B). Real-time feedback on performance is given by a running timer, an alarm signal combined with a written warning on the arthroscopic image when a force threshold level is exceeded**.** The trainees' overall performance after completion of an exercise is presented in the form of a spider graph.

During the validation tests, additional equipment was used to document the participant's performance. Recording equipment was set up

to document the task times independently of the simulator. Simultaneous digital recordings were acquired from the arthroscope and a separate camera (digital CCD camera, 21CW Sony CCD, Tokyo, Japan).

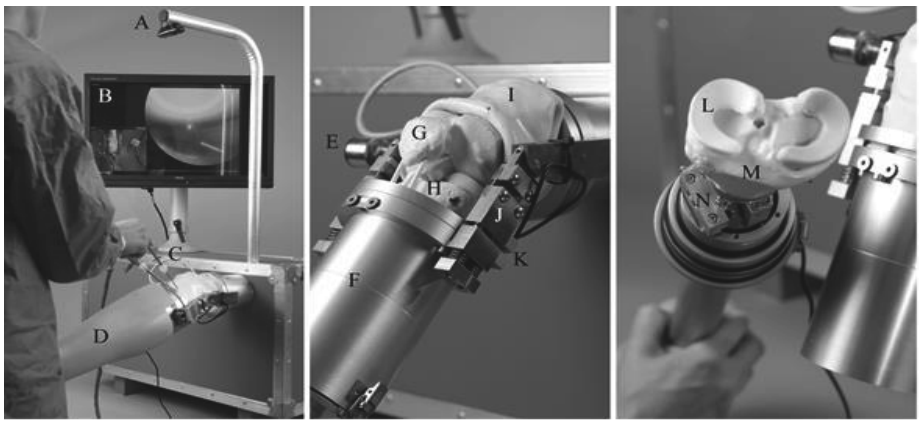


Figure 1. Motion is measured in a two-dimensional plane using a webcam (**A**) and yellow-coloured markers (**C**). A graphical user interface was designed leading trainees through a range of exercises, indicating progress in training by means of spider graph and giving real-time feedback on performance during exercises (**B**).The outer appearance of the lower leg was made from a dummy leg of a mannequin (**D**). Two magnetic hall sensors were positioned in line with the springs (**K**). The inner construction of the lower leg was designed such that it allowed complete and easy removal of the inner metal construction to which the tibia plateau (M) and menisci (L) were connected by unlocking one pin positioned in the heel and two clips (**F**).A patellar bone (**G**) was added to the intra-articular structures as well as cruciate ligaments (H). The skin contains two prefabricated portals that are routinely used during knee arthroscopy (**I**). A special hinge was designed (**J**) to allow flexion-extension, which as measure by an angular potentiometer (E) and to allow joint stressing. One force sensor was connected to the tibia plateau and one force sensor to the femoral condyles to measure instrument-tissue interaction (**N**)

Participants

All surgeons, residents and researchers in our department that were eligible for participation in the study were recruited. Thirty-one participants were divided into three groups having different levels of arthroscopic experience: beginners who had never performed an arthroscopic procedure, intermediates who had performed up to 59 arthroscopies, and experts who had performed 60 or more arthroscopies.

This boundary level of 60 arthroscopies was based on the average opinion of fellowship directors who were asked to estimate the number of operations that should be performed to allow a trainee to perform unsupervised meniscectomies[42].

Study design

To test for face and construct validity, a previously used protocol was applied[35]. Firstly, participants answered questions regarding general information (Fig. 2) and the outer appearance of the simulator for face validity. Secondly, the construct validity test was performed using normal arthroscopic equipment (Smith and Nephew, Andover Massachusetts USA). Finally, face validity, educational value and user friendliness were further determined by giving the participants exercises that were representative for the simulator's specific features and by subsequently completing the questionnaire[35].

The assessment of construct validity was based on one basic navigation task that was timed[33, 35]. If participants failed to complete the task within ten minutes, they continued to the next part of the study protocol[33]. The task time was determined with a separate video recording of the simulator monitor that shows the intra-articular joint. Accurate determination of the task time was ensured, since the video recordings had a frame rate of 25 images per second and allowed replay frame by frame. This resulted in an accuracy of 0.04 seconds. In the subsequent part of the protocol, the participants were asked to perform two additional exercises representative for the simulator's specific features. The first exercise was to perform a medial or lateral meniscectomy. The assignment was to cut the posterior horn of the medial or lateral meniscus and acquire an impression of tissue when cutting. The second exercise was to perform a

guided navigation task. For this task, three landmarks had to be identified by probing; medial femoral condyle, ACL and posterior horn of the lateral meniscus. Probing was preceded by textual instructions presented by augmented reality on top of the intra-articular view of the arthroscope. The two exercises both had to be performed once; no task time or other parameters were recorded. However, the participants were pointed out to the user interface and the manner for presenting feedback on performance.

Questions were filled out concerning face validity of the outer appearance, the intra-articular joint, the instruments and instrument handling. Educational Value I comprised the variation and level of exercises; Educational Value II indicated to what extent the content of the simulator is a good way to prepare for real-life arthroscopic operations. User Friendliness I, covered questions concerning the quality of the instructions given by the simulator and the presentation of the performance; User Friendliness II (Table 1) indicated whether the participants needed a manual before operating the simulator[35].

The questions were answered using a 10-point numerical rating scale (NRS) (i.e. 0 = completely unrealistic and 10 = completely realistic) or a dichotomous scale requiring a yes/no answer[35]. A ‘‘not applicable (N/A)’’ option, could be used solely by beginners. Only the answers from the expert and intermediate groups were used on simulator realism and Educational Value I. These questions required prior knowledge of the real-life arthroscopic situation. A value of 7 or greater was considered sufficient. Participants were also asked to give free text suggestions for improvements[35].

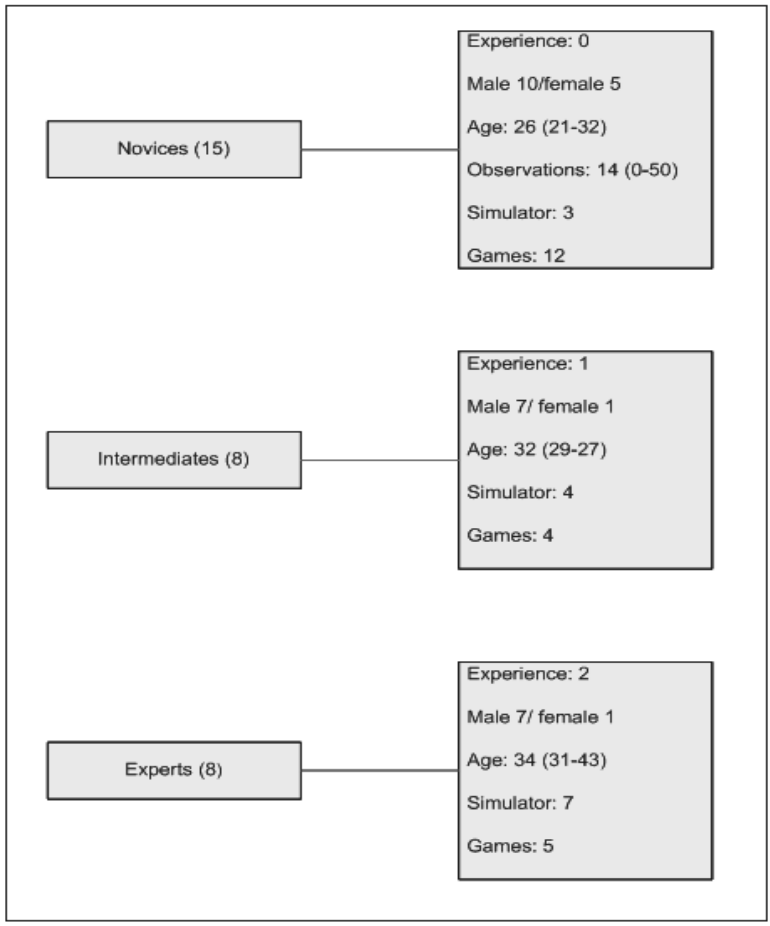


Figure 2. This figure shows the participant population per experience group. The age in years and the number of attended arthroscopies (''Observation'') are expressed as median with range. Also the number of participants who previously had used a simulator (''Simulator'') or had experience in playing computer games (''Games'') is shown

Table 1: The median scores for face validity of the simulator (rows 1-3), Educational Value I and User Friendliness I

Face validity:	Median (range):
Outer appearance	8.7 (6-9.3)
Intra-articular joint	7.5 (4.8-8.8)
Surgical instruments	8.3 (6-10)
User friendliness I	7.3 (1.8-9.3)
Educational Value I	7 (2-8)

Statistical analysis

One observer determined all navigation task times using the recorded arthroscopic view by consecutive determination of each identified landmark. This way, the proper sequence in the landmark identification protocol was guaranteed. One complete task time was calculated as the summation of the duration between all landmarks of one trial. Statistical analysis was performed using SPSS 20 (SPSS Inc., Chicago, IL, USA). The presence of normal distributions of task times was assessed by Shapiro Wilk tests. As the data were not normally distributed, non-parametric tests were performed. Construct validity was assessed with the Kruskal-Wallis test by calculation of overall significant differences in task time between the three groups for each of the five task trials. The significance level was adjusted for multiple comparisons with the Bonferroni-Holm procedure (alpha = 0.05)[43]. Mann-Whitney U tests were used for pair-wise comparisons to highlight significant differences. The results of the three separate aspects of face validity of the simulator, Educational Value I and User Friendliness I were expressed as mean summary scores of the corresponding grouped questions. As these results were not normally distributed they were expressed as median (minimum-maximum). The dichotomous questions (Educational Value II and User Friendliness II) are presented as frequencies and percentages (%)[35]. Data from previous research on simulators[35] for which the same study design was used were applied to determine if PASSPORT V2 shows improved realistic feel. Therefore, the results of the 'intra-articular joint' -and 'instruments' aspects of face validity were compared with the Kruskal-Wallis test and the Mann Withney U test ($p<0.5$).

Results

Two out of fifteen beginners completed three out of five trials within the time limit. Three out of fifteen beginners and one out of eight intermediates completed four out of five trials. The median task time of the first trial was 137 seconds (range, 57-240 seconds) for the beginners, 99 seconds (range, 37-166 seconds) for the intermediates, and 69 seconds for the experts (range, 52-109 seconds). The median task time for the fifth trial was 55 seconds (range, 17-139 seconds) for the beginners, 33 seconds (range, 17-59 seconds) for the intermediates, and 26 seconds (range, 14-52 seconds) for the experts (Fig. 3). The beginners were significantly slower than the experts in completing all five trials. The task times of the beginners and the intermediates were not significantly different for the first, third and fifth trial. The second and fourth trials were performed significantly faster by the intermediates. The task times of the intermediates and the experts were not significantly different for all trials (Fig. 3).

The outer appearance of any simulator was indicated as important by 60% of the participants. Ninety per cent of the participants indicated the outer appearance of this simulator as sufficient. All median values of the summed scores for face validity, Educational Value I and User Friendliness I were 7 or higher (Tables 1, 2). Comparison of the face validity with the two previously evaluated simulators showed that the 'intra-articular joint' and 'instruments' aspects of face validity of the PASSPORT were both significantly higher ($p<0.001$) compared to simulator 1[35]. Compared to simulator 2, the 'instruments' aspect of the PASSPORT was significantly higher ($p<0.001$). The median sum score for Educational Value I was 4 out of 5 (range 1-4). All intermediates and experts indicated that the simulator is a good way to prepare for

arthroscopic operations (Educational Value II). Ninety percent of the participants did not feel the need to read the manual (User Friendliness II) before operating the simulator. Sixty-five percent of the participants indicated that the simulator is most valuable as a training modality during the first year of the residency curriculum. Main advantages of the simulator were the realistic arthroscopic view and probing, the realistic performance of a meniscectomy and the possibility to train important aspects of a knee arthroscopy, especially navigation and hand-eye coordination. Participants indicated the presence of realistic tactile feedback as the biggest asset of the simulator (fourteen participants out of 31), and considered it an essential condition to imitate clinical practice. Most important suggestions for improvements included: fine-tuning of the appearance of the warning on the arthroscopic image when the threshold level of the force parameters is exceeded (seven participants), more realistic material for the cruciate ligaments and the menisci (six participants) and a more gradual joint stressing (four participants). Points of criticism were the absence of the popliteal ligament, and the fact that the performance spider graph showed software bugs leading to incorrect presentation of the performance results.

Table 2: The median scores for individual questions for face validity

Face validity intra-articular:	Median (range)	Face validity Instruments	Median (range)
Intra-articular anatomy	7 (5-9)	Instruments visual	8 (5-10)
Texture structures	7 (4-8)	Instruments motion	9 (5-10)
Color structures	7 (5-9)	Instruments probing	7 (5-9)
Size intra-articular joint space	8 (5-10)	Instruments cutting	7 (0-9)
Arthroscopic image	7,5 (4-9)		

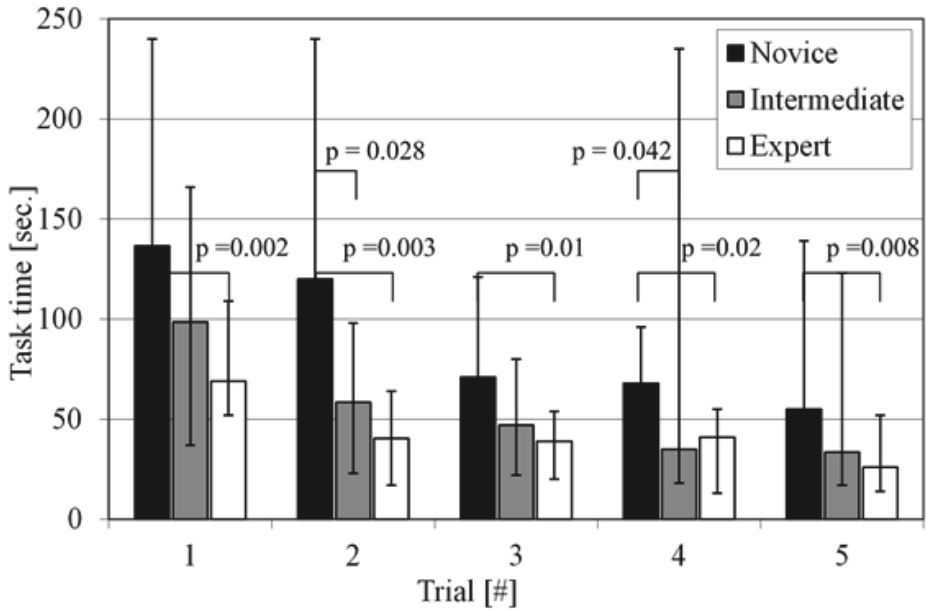


Figure 3. Median task time per trial for each experience group

Discussion

The PASSPORT V2 prototype was found to show full face validity as all median scores were 7 or higher (Tables 1, 2). Construct validity was demonstrated for every trial, since the experts were significantly faster than the beginners (Fig. 3). PASSPORT V2 was favored by experts and residents on the intra-articular joint and instruments aspects of face validity compared to two previously tested simulators[35].

Differences in task time between beginners and intermediates were only partly significant, and no significant difference was found between intermediates and experts. This might be caused by the fact that the intermediate group was the most heterogeneous group with respect level of experience. As a consequence, relatively large differences existed, contributing to a higher variance in task times (Fig. 3). Behavioral factors could also have been of influence: most beginners were enthusiastic and

motivated to learn new skills, while intermediates tended to be more competitive, which could have contributed to being less cautious. The experts were preoccupied with being accurate and cautious in treating the tissue rather than trying to perform the task as fast as possible. It was suggested previously that the influence of these behavioral factors affect the outcome[26, 33].

Although 40% of the participants did not care about the outer appearance of simulators in general, most participants (90%) did indicate that the outer appearance of the PASSPORT is sufficient (Table 1). The arthroscopic view inside the PASSPORT knee phantom was considered sufficiently realistic by the intermediates and experts (Tables 1, 2), which indicated the right direction in improving realism. Still, room for improvement was pointed out: the material and fixation of the cruciate ligaments as well as the menisci and more gradual sense of stressing. The latter has been taken into account by replacing the springs by a different set. All participants considered training on this simulator suitable to prepare for arthroscopic operations. Educational Value I and II and User Friendliness I and II were considered sufficient. Instructions given by the user interface were perceived as clear and intuitive by all participants. Highly appreciated was the presence of adequate tactile feedback, which offered realistic navigation, probing and cutting in the intra-articular joint space. Tactile feedback is considered essential to provide a sense of realism[19, 44, 45] and the participants confirmed this and indicated that PASSPORT V2 does offer this[29, 31]. Overall, all participants agreed that apart from some adjustments the PASSPORT V2 can evolve in a valuable training tool in the first year of the residency curriculum.

A limitation of this study was the relatively small number of participants in each experience group. This could have been the reason

that significant differences between the beginner and intermediate group and between the intermediate and expert groups were not detected. However, earlier mentioned factors (the heterogeneity of the intermediate group and behavioural factors) are more likely to have contributed to the absence of significant differences. Moreover, a protocol was applied that was used previously which strengthens the results of the current study[35]. Another limitation of our study is that only the task time as parameter was recorded as prescribed in the protocol.

The clinical relevance of this study is the fact that evidence of validation of a simulator is necessary before it is used in practice to train residents. Moreover, the PASSPORT V2 has overcome critical drawbacks of virtual reality simulators, most necessarily in its realistic force feedback. The simulator has also proven to be an improvement compared to the PASSPORT V1 in its full integration of registration devices and a reasonable intuitive user interface[35]. Future plans are to perform test with irrigation and the generation of bleedings at random moments during an exercise and testing of adequate tissue manipulation.

Conclusion

The results support the face and construct validity of the PASSPORT V2 and this simulator is recognized as a valuable training modality by experienced surgeon and residents. Moreover, the PASSPORT V2 shows enhanced force feedback, which is considered a key feature for arthroscopic simulators.

Acknowledgement

We thank the Marti-Keuning Eckhart Foundation (Lunteren, The Netherlands) that funded this research. The development of ForMoST v2 was funded by STW valorisation grant 1, Utrecht, The Netherlands. The special hinge was designed by the company MediSHield BV (Delft, Netherlands). The authors are grateful to Tjeerd Dudink, Arne Hofmann, Erwin Langelaar, Mathijs de Wit, Arjan van Dijk, Daisy Feng, Pieter Bas Wulms, Tijmen Blankevoort, Maarten van Zomeren, Andries Oort, John Dukker, Freek van Delft and Willem Nerkens for their contribution to the development of the PASSPORT V2.

References

[1] Escoto A, Trejos AL, Naish MD, Patel RV, Lebel ME: Force sensing-based simulator for arthroscopic skills assessment in orthopaedic knee surgery Stud Health. Technol Inform 2012, 173:129-135.

[2] Bridges M, Diamond DL: The financial impact of teaching surgical residents in the operating room. Am J Surg 1999, 177:28-32.

[3] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Br 2008, 90:494-499.

[4] Slade Shantz JA, Leiter JR, Gottschalk T, MacDonald PB: The internal validity of arthroscopic simulators and their effectiveness in arthroscopic education. Knee Surg Sports Traumatol Arthrosc 2014, 22:33-40.

[5] Gomoll AH, Pappas G, Forsythe B, Warner JJ: Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. Am J Sports Med 2008, 36:1139-1142.

[6] Reznek M, Harter P, Krummel T: Virtual Reality and Simulation: Training the Future Emergency. Physician Virtual Real Simul 2002, 9:78-87.

[7] Modi CS, Morris G, Mukherjee R: Computer-simulation training for knee and shoulder arthroscopic surgery. Arthroscopy 2010, 26:832-840.

[8] van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J: Objective assessment of technical surgical skills. Br J Surg 2010, 97:972-987.

[9] Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M, Hutchison C: Assessment of technical skills transfer from the bench training model to the human model. Am J Surg 1999, 177:167-170.

[10] Brehmer M, Tolley D: European Urology Validation of a Bench Model for Endoscopic Surgery in the Upper Urinary Tract. Eur Urol 2002, 42:175-180.

[11] Brehmer M, Swartz R: Training on Bench Models Improves Dexterity in Ureteroscopy. Eur Urol 2005, 48:458-463.

[12] Cannon WD, Eckhoff DG, Garrett Jr. WE, Hunter RE, Sweeney HJ: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res 2006, 442:21-29.

[13] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35: 883-888.

[14] Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. Br J Surg 2004, 91:146-150.

[15] Jacomides L, Ogan K, Cadeddu JA, Pearle MS: Use of a virtual reality simulator for ureteroscopy training. J Urol 2004, 171:320-323.

[16] Knoll T, Trojan L, Haecker A, Alken P, Michel MS: Validation of Computer-Based Training in Ureterorenoscopy. BJU Int 2005, 95:1276-1279.

[17] Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD: A Novel Approach to Endourological Training: Training at the Surgical Skills Center. J Urol 2001, 166:1261-1266.

[18] Matsumoto ED, Pace KT, D'A Honey RJ: Virtual Reality Ureteroscopy Simulator as a Valid Tool for Assessing Endourological Skills. Int J Urol 2006, 13:896-901.

[19] McCarthy AD, oody ML, Waterworth AR, Bickerstaff DR: Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res 2006, 442: 13-20.

[20] Michelson J: Simulation in orthopaedic education: an overview of theory and practice. J Bone Joint Surg 2006, 88:1405-1411.

[21] Naik VN, Matsumoto ED, Houston PL, Hamstra SJ, Yeung RY, Mallon JS, Martire TM: Fiberoptic orotracheal intubation on anesthetized patients: do manipulation skills learned on a simple model transfer into the operating room? Anesthesiology 2001, 95:343-348.

[22] Ogan K, Jacomides L, Shulman MJ, Roehrborn CG, Cadeddu J, Pearle MS: Virtual Ureteroscopy Predicts Ureteroscopic Proficiency of Medical Students on a Cadaver. J Urol 2004, 172:667-671.

[23] Oostema JA, Abdel MP, Gould JC: Time-efficient laparoscopic skills assessment using an augmented-reality simulator. Surg Endosc 2008, 22:2621-2624.

[24] Pedowitz RA, Esch J, Snyder S: Evaluation of a virtual reality simulator for arthroscopy skills development. Arthroscopy 2002, 18(6):E29.

[25] Seymour N, Gallagher AG, Roman, SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM: Virtual Reality Training Improves Operating Room Performance Results of a Randomized, Double-Blinded Study. Annals of surgery 2002, 236:458-464.

[26] Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL: Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. J Shoulder Elbow Surg 2004, 13:196-205.

[27] Watterson JD, Beiko DT, Kuan JK, Denstedt JD: Randomized Prospective Blinded Study Validating Acquistion of Ureteroscopy Skills Using Computer Based Virtual Reality Endourological Simulator. J Urol 2002, 168:1928-1932.

[28] White MA, Dehaan AP, Stephens DD, Maes A, Maatman TJ: Validation of a High Fidelity Adult Ureteroscopy and Renoscopy Simulator. J Urol 2010, 183:673-677.

[29] Moody L, Waterworth A, McCarthy AD, Harley P, Smallwood R: The feasibility of a mixed reality surgical training environment. Virtual Reality 2008, 12:77-86.

[30] Safir O, Dubrowski A, Mirksy C, Backstein D, Carnahan H: What skills should simulation training in arthroscopy teach residents? Int J Cars 2008, 3:433-437.

[31] Zivanovic A, Dibble E, Davies B, Moody L, Waterworth A: Engineering requirements for a haptic simulator for knee arthroscopy training. Stud Health Technol Inform 2003, 94: 413-418.

[32] Okamura A: Haptic feedback in robot-assisted minimally invasive surgery Curr Opin Urol 2009, 19(2):102-107.

[33] Tuijthof GJ, van Sterkenburg MN, Sierevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills Knee Surg Sports Traumatol Arthrosc 2010, 18:218-224.

[34] Leong JJ, Leff DR, Das A, Aggarwal R, Reilly P, Atkinson HD, Emery RJ, Darzi AW: Validation of orthopaedic bench models for trauma surgery. J Bone Joint Surg Br 2008, 90:958-965.

[35] Tuijthof GJ, Visser P, Sierevelt IN, van Dijk CN, Kerkhoffs GM: Does perception of usefulness of arthroscopic simulators differ with levels of experience? Clin Orthop Relat Res 2011, 469:1701-1708.

[36] Tuijthof GJM, HoremanT: Training facility, surgical instruments and artificial knee with an upper limb and a lower limb for simulation and training of arthroscopic surgical techniques. 2011.

[37] Stunt JJ, Wulms PH, Kerkhoffs GM, Sierevelt IN, Schafroth MU,Tuijthof GJ: Variation in joint stressing magnitudes during knee arthroscopy. Knee Surg Sports Traumatol Arthrosc 2014, 22:1529-1535.

[38] Verdaasdonk EG, Stassen LP, Monteny LJ, Dankelman J: Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. Surg Endosc 2006, 20:511-518.

[39] Verdaasdonk EG, Stassen LP, Schijven MP, Dankelman J: Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Surg Endosc 2007, 21:1406-1412.

[40] Verdaasdonk EG, Stassen LP, van Wijk RP, Dankelman J: The influence of different training schedules on the learning of psychomotor skills for endoscopic surgery. Surg Endosc 2007, 21:214-219.

[41] Horeman T, Rodrigues SP, Jansen FW, Dankelman J, van den Dobbelsteen JJ: Force measurement platform for training and assessment of laparoscopic skills. Surg Endosc 2010, 24:3102-3108.

[42] O'Neill PJ, Cosgarea AJ, Freedman JA, Queale WS, McFarland EG: Arthroscopic proficiency: a survey of orthopaedic sports medicine fellowship

directors and orthopaedic surgery department chairs. Arthroscopy 2002, 18:795-800.

[43] Holm S: A simple sequentially rejective multiple test procedure. Scand J Statist 1979, 6:65-70.

[44] Müller W, Bockholt U: The virtual reality arthroscopy training simulator. Stud Health Technol Inform 1998, 50:13-19.

[45] Torkington J, Smith SG, Rees BI, Darzi A: The role of simulation in surgical training. Ann R Coll Surg Engl 2000, 82:88-94.

CHAPTER 5

Face and construct validation of Virtamed Virtual Arthroscopic knee simulator

J.J. Stunt, G.M.M.J. Kerkhoffs, C.N. van Dijk, G.J.M. Tuijthof

Knee Surg Sports Traumatol Arthrosc (2015) 23: 3435-3442

Abstract

Purpose: Virtual reality simulator-training has become important for acquiring arthroscopic skills. A new simulator for knee arthroscopy ArthroS™ has been developed. The purpose of this study is to demonstrate face and construct validity, executed according to a protocol used previously to validate arthroscopic simulators.

Methods: Twenty-seven participants were divided into three groups having different levels of arthroscopic experience. Participants answered questions regarding general information and the outer appearance of the simulator for face validity. Construct validity was assessed with one standardized navigation task. Face validity, educational value and user friendliness were further determined by giving participants three exercises and by asking them to fill out the questionnaire.

Results: Construct validity was demonstrated between experts and beginners. Median task times were not significantly different for all repetitions between novices and intermediates, and between intermediates and experts. Median face validity was 8.3 for the outer appearance, 6.5 for the intra-articular joint and 4.7 for surgical instruments. Educational value and User-friendliness were perceived as non-satisfactory, especially because of the lack of tactile feedback.

Conclusion: The ArthroS™ demonstrated construct validity between novices and experts, but did not demonstrate full face validity. Future improvements should be mainly focused on the development of tactile feedback. It is necessary that a newly presented simulator is validated to prove it actually contributes to proficiency of skills.

Introduction

Arthroscopic surgery has developed into an important operative therapy. As arthroscopic skills are complex, learning these skills is challenging and time consuming[1]. Traditionally, arthroscopic training is performed in the operating room using a one-to-one apprentice model, where the resident is permitted to perform actions under supervision. This is not an ideal learning environment. The resident is not able to make mistakes and learn from them, because these mistakes would compromise patient safety[2-4]. Inspired by the aerospace industry, virtual reality simulation has been deployed successfully in the medical world on a variety of medical skills[5, 6]. Due to reductions in training time and increasing demands for patient safety and quality control, virtual reality simulators have also been introduced in the education of orthopaedic residents[3, 7, 8]. The potential advantages of arthroscopic knee simulators has been demonstrated repeatedly[1, 3, 4, 7, 9, 10]. However, objective validation of individual simulators that are new to the market remains necessary to confirm their contribution to the improvement of skills.

A new virtual reality simulator for knee arthroscopy has been launched (ArthroS™, VirtaMed AG (www.virtamed.com). This simulator has been developed through a joint effort of medical experts from the Balgrist University Hospital and the ETH Zurich. The aim of this new simulator is the presentation of a realistic virtual training environment, since other virtual reality simulators evaluated for training of knee arthroscopy[11-20], only showed a marginal level of realism[21-24]. Similar to other simulators, the simulator offers training of diagnostic and therapeutic arthroscopy and objective performance feedback. However, the method to determine collision of the instrument tools with the anatomic structures is performed differently, which is meant to enhance

the sense of realistic tactile feedback. Additionally, irrigation flow and bleedings is simulated to improve realism. The purpose of this study is to determine face and construct validity of this simulator. To allow comparison, this will be performed according to a protocol that has previously been used for the validation of arthroscopic knee simulators[4].

Materials and Methods

Simulator characteristics

The simulator has an original arthroscope and original instruments (palpation hook, grasper, cutting punch and shaver) that were modified to connect them to the virtual environment of this simulator. The arthroscope features in- and outlet valves for fluid handling and three virtual cameras (0º, 30º and 70º) including a focus wheel. The simulator provides a patient knee model and a high-end PC with touch-screen. The VirtaMed ArthroS™ training software consists of 6 fully guided procedures for basic skill training, four virtual patients for diagnostic arthroscopy, eight virtual patients for various surgical operations and courses designed for beginners, intermediate trainees and advanced arthroscopists. It is possible to display an anatomical model of the knee in which the real-time orientation of the instruments is shown. Registration devices are integrated to provide objective feedback reports and registration of training sessions.

Table 1: The median scores for face validity of the simulator (rows 1-3), Educational Value I and User Friendliness I

Face validity:	**Median (range):**
Outer appearance	8.3 (5.3-9.7)
Intra-articular joint	6.5 (4.2-8.5)
Surgical instruments	4.7 (3.3-6.7)
User friendliness	7.8 (4.5-8.8)

Participants

Twenty-seven participants were recruited to perform the validity test. To keep in line with the protocol, the same strategy was applied in dividing the participants in three groups with different levels of arthroscopic experience: beginners who had never performed an arthroscopic procedure, intermediates who had performed up to 59 arthroscopies and experts who had performed more than 60 arthroscopies (Table 1). This boundary level of 60 arthroscopies was based on the average opinion of fellowship directors who were asked to estimate the number of operations that should be performed before allowing a trainee to perform unsupervised meniscectomies[25]. It was investigated whether the simulator could discriminate between the three groups.

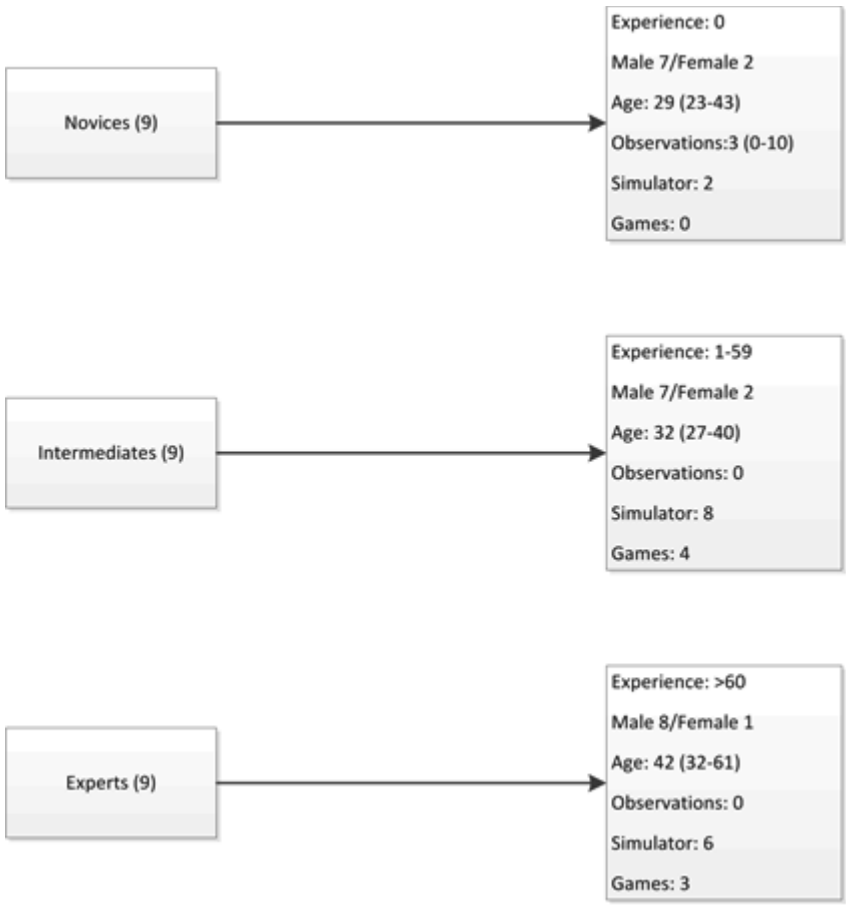


Figure 1: The participant population per experience group. The age in years and the number of attended arthroscopies (''Observation'') are expressed as median with range. Also the number of participants who previously had used a simulator (''Simulator'') or had experience in playing computer games (''Games'') is shown

Study design

To ensure a similar level of familiarity with using the simulator, the participants were not allowed to use the simulator before the experiment. A previously developed test protocol was used to determine face and construct validity[4]. Firstly, participants were asked to answer questions regarding general information (Fig. 1) and to provide their opinion on the outer appearance of the simulator for face validity[4]. The assessment of

construct validity was performed on one navigation task[26], that was to be repeated five times. The time to complete all repetitions was set at 10 minutes[4]. Task time was used as an outcome measure, as it enables objective comparison between the experience groups[1]. The task time was determined with a separate video recording of the simulator's monitor in which the intra-articular joint is presented. This enabled verification of the correct sequence of probing the nine anatomic landmarks. Accurate determination of the task time was ensured, since the video recordings had a frame rate of 25 images per second (equal to accuracy an of 0.04 seconds).

Subsequently, three exercises representative for the capability of this particular simulator were performed. The first task was a triangulation task to train hand-eye coordination. The second task was a guided navigation task for a complete inspection of the knee (Fig. 2). The third task was a medial meniscectomy. During the meniscectomy it was possible to perform joint irrigation. For these three exercises, no task time was recorded, but rather the participants' impression of the capabilities of the simulator. After every task, an objective feedback report, and a before –and after picture were shown to demonstrate the way in which feedback was provided by the simulator.

After the three tasks, the participants filled out the remainder of the questionnaire[4]. The questionnaire comprised questions concerning face validity, educational value, and user friendliness. Face Validity was determined based on several aspects of the simulator (Table 1). Educational Value I concerned the variation and level of exercises; Educational Value II indicated to what extent the simulator serves as a good way to prepare for real-life arthroscopic operations. User Friendliness I comprised questions concerning the quality of the

instructions given by the simulator and the presentation of the performance and results; User friendliness II indicated whether the participants needed a manual before operating the simulator[4].

Questions were answered using a 10-point numerical rating scale (NRS) (e.g., 0 = completely unrealistic and 10 = completely realistic) or dichotomous requiring a yes/no answer[4]. A value of 7 or greater was considered sufficient. Questions featuring a ‘‘not applicable (N/A)’’ answer option, could be used solely by beginners. Furthermore, only the answers from the expert and intermediate groups were used to determine Face Validity and Educational Value I. Participants were able to provide free text suggestions for improvement of the simulator.

IRB approval

Academic Medical Centre, Amsterdam; reference number:
W13_262 # 1.17.0326.
The Medical Research Involving Human Subjects act does not apply to the current study and an official approval of this study by the Medical Ethical Commission is not required.

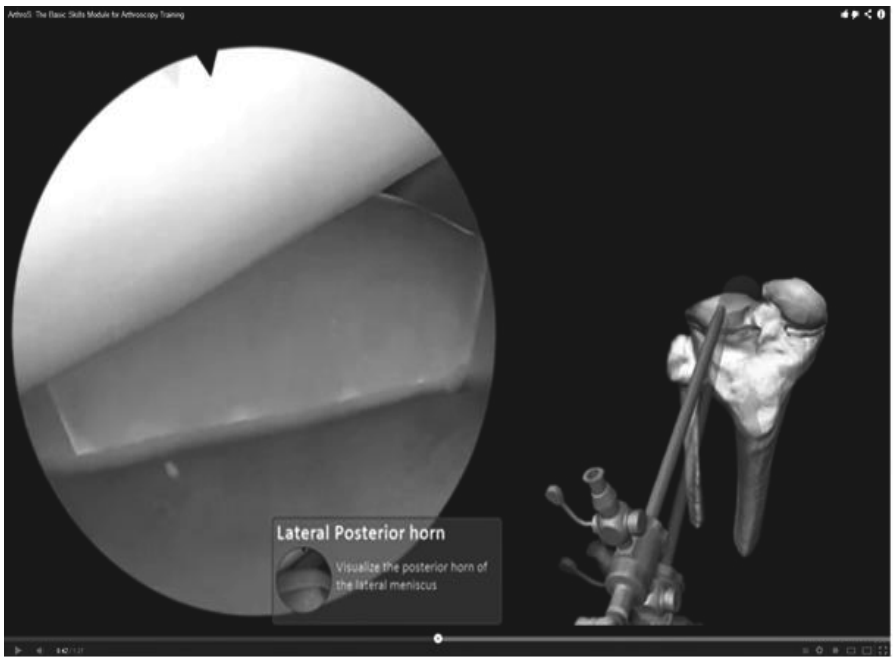


Figure 2: A ghost tool on the monitor demonstrates the correct position of the scope to visualize the indicated structures within the knee joint

Statistical analysis

Data processing and statistical analysis was performed on the basis of the study protocol according to a previous study[4]. One observer determined all navigation task times using the recorded arthroscopic view by consecutive determination of each identified landmark. The summation of the duration to probe all landmarks of one trial was considered as one complete task time. Statistical analysis was performed using SPSS 20 (SPSS Inc., Chicago, IL, USA). The Shapiro Wilk test demonstrated a skewed distribution. Therefore, nonparametric tests were performed, and values were presented as median (minimum-maximum). Construct validity was assessed by calculation of significant statistical differences in the task time for each trial between the three groups with the nonparametric Kruskal Wallis-test and Mann–Whitney U tests to highlight

significant differences. The significance level was adjusted for multiple comparisons with the Bonferroni-Holm procedure (alpha = 0.05). The results of Face Validity, Educational Value I and User Friendliness I were expressed as median summary scores of the corresponding questions.

Results

One beginner completed only one task trial within the time limit. All other participants completed the five trials. The median task times and ranges of each trial are represented in Fig. 3. The median task time for the fifth trial was 41 seconds (3-50 seconds) for the beginners, 35 (18-120 seconds) for the intermediates, and 31 seconds (17-42 seconds) for the experts (Fig. 3).

Task times of the beginners and the intermediates did not show significant differences for all repetitions (n.s.), nor did the task times of the intermediates and the experts (n.s.). The task times of the beginners were significantly slower than those of the experts for all trials (Fig. 3).

The outer appearance was indicated as satisfactory by the experts and intermediates (Table 1 median > 7). The intra-articular face validity and instrument face validity were not sufficient (Table 1 median < 7), because the texture of the structures, the motion of instruments, and tissue probing and cutting were perceived as not realistic (Table 2). Related to these results were the remarks written by respondents: presence of a small delay between the motion of the instruments and the virtual image (fourteen participants) and markers were sometimes placed in the femur condyle (four participants).

The quality of instructions was perceived well (Table 1, User Friendliness I). Seventy-one percent of the participants did not feel the need to read the manual (User Friendliness II). Fifty-nine percent of the participants indicated that the simulator is most valuable as a training

modality during the first year of the residency curriculum. The level of exercises was 2 out of 5 (range 1-5) (Educational Value I). However, the variation was judged adequate. In fact, the extensive training program, which enables to practice a wide range of procedures, especially hand-eye coordination, was considered an important asset of the simulator. Fifty percent of the intermediates and experts agreed that training on this simulator is a good way to prepare for arthroscopic operations (Educational Value II).

Important suggestions for improvement included: height adjustment of the lower leg (three participants), replacement of the tibia bone by a complete lower leg for better varus-valgus stressing (ten participants), more realistic use and movement of the instruments (six participants) and the improvement of tactile feedback (eighteen participants).

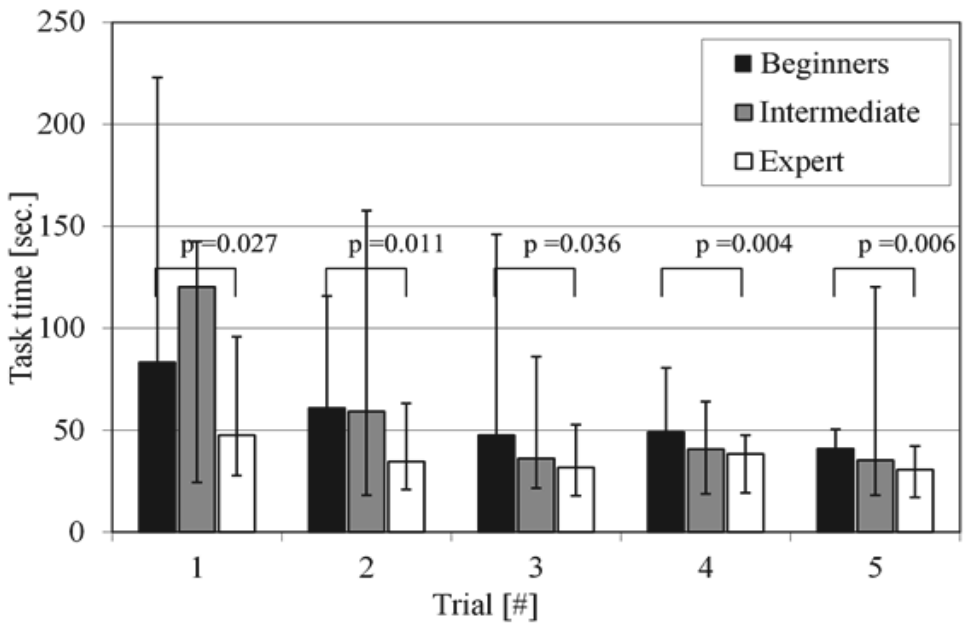


Figure 3: Median task times [minimum-maximum] of novices, intermediates and experts of each repetition

Discussion

The most important finding of the present study was that the simulator showed construct validity between beginners and experts for the navigation task (Fig. 3) and that face validity was partly achieved. No significant differences were found between trials of the beginners and the intermediates and between the intermediates and the experts ($p>0.05$). Analyzing the task times, different factors could have contributed to this. Overall all task times were relatively fast, which could indicate that with this simulator the navigation task was fairly easy to perform. As all participants needed to become familiar with the simulator including the experts, learning pace might have been the same for all experience levels. Another noticeable result is the relative large variance of the task times in the intermediate group. As this was the most heterogeneous group in terms of skills levels - some intermediate participants performed no more than few knee arthroscopies, while others performed almost sixty – this explains the large variance. However, some studies did show significant differences between experience groups[1]. These studies both stratified their subjects into three groups (beginner, intermediate and experts). In the study of Pedowitz et al, larger numbers of participants per experience group (35, 22 and 21 respectively) were included, which could have increased the chance to detect significant differences. However, they did not classify their participants based on a certain number of performed arthroscopies[1]. The study of Srivastava et al is more comparable to the current study, since they included a similar number of participants, and retained a comparable expert boundary (50 arthroscopies)[27]. As there was no correlation between task times and experience in playing computer games, it is not probable that gaming experience contributed to the lack of significant differences.

Almost all participants indicated that the outer appearance of the simulator was sufficient. All essential intra-articular anatomical structures were present and were considered realistic in terms of their size. However, stressing of the knee, and the color and the texture of the structures were regarded less realistic, which is remarkable as this virtual reality simulator makes use of high-end graphics and is the first to simulate irrigation. The latter is meant to provide an additional sense of realism as was found in Tuijthof et al.[26]. The joint stressing issue can be easily solved by incorporating the suggestion of the participants to extend the rather short partial lower leg by a full sized one. Although a new manner to simulate tissue probing and cutting was implemented in the system, the participants felt that the haptic feedback they experienced did not adequately imitate clinical practice. This was reflected in the low scores for surgical instruments use (Table 1, 2) and the perceived educational value.

Questionnaire results showed that tissue probing and cutting, texture of the structures and irrigation were perceived as not realistic. As a consequence, proper training of joint inspection and therapeutic interventions was not considered feasible (Educational Value I). Eighteen participants explicitly mentioned the absence of realistic tactile feedback as a limitation in using the simulator as a training modality. Detailed analysis revealed that this might have been caused by an offset of the system's calibration that connects the real to the virtual world, as the markers of one of the tasks were occasionally positioned inside the femoral condyle. The importance of realistic tactile feedback has been raised earlier[28-31], and the lack of (realistic) feedback has been considered a limitation[18]. The current and previous studies demonstrate that simulators without realistic tactile feedback are perceived as less appropriate for training arthroscopic skills[4, 15]. Simulators that do give

realistic (partly) tactile feedback are considered more suitable for training and have shown to improve arthroscopic skills in the operating room[2, 32, 33].

When comparing the results of this virtual reality simulator to other virtual reality simulators that have been tested in literature, the simulator offers similar results as the GMV simulator in offering acceptable face validity, and a wide variety in exercises via a user friendly interface[2, 4].

A limitation of this study was the relatively small number of participants in each experience group. It is possible that significant differences in the navigation trials were not detected. For logistic reasons this was the maximum number of participants that we could include. Besides, the heterogeneity of the intermediate group is likely to also have contributed to the absence of significant differences. Another limitation is that we only tested construct validity for one task, instead of testing the complete curriculum of exercises. Within the set time frame of 30 minutes this was not possible, but due to this time limit we were able to include a sufficient number of experts.

Obtaining arthroscopic skills is challenging. Especially at the start of their learning curve it is important that trainees are able to make mistakes that do not compromise patient safety. Therefore, traditional arthroscopic training in the operating room is not ideal[2-4]. Simulators can play a valuable role in training residents, but this has to be verified through objective validation protocols for each individual simulator. This study is clinically relevant as the findings provide important information for educators that intend to use this simulator to train residents; it is necessary that a newly presented simulator demonstrates validity. The strength of this study is the application of a previously introduced study design that allows relative comparison in an objective manner and highlights areas of

improvement. Overall, the simulator was considered a reasonable preparation for real-life arthroscopy with as its main advantages the large variety of exercises, the extensive theoretical section and the realistic intra-articular anatomy. However, essential improvements are necessary which mainly focus on the development of realistic tactile feedback.

Conclusion

The simulator demonstrated construct validity between beginners and experts. Face validity was not fully achieved. The most important shortcoming of the simulator is the lack of sufficient tactile feedback, which participants considered essential for proper arthroscopic training. According to most participants the simulator ArthroS™ has potential to become a valuable training modality.

References

[1] Pedowitz RA, Esch J, Snyder S: Evaluation of a virtual reality simulator for arthroscopy skills development. Arthroscopy 2002, 18:E29.

[2] Bayona S, Fernandez-Arroyo JM, Martýn I, Bayona P: Assessment study of insightARTHRO VR arthroscopy virtual training simulator: face, content, and construct validities. J of Robo Surg 2008, 2:151-158.

[3] Gomoll AH, Pappas G, Forsythe B, Warner JJ: Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. Am J Sports Med 2008, 36:1139-1142.

[4] Tuijthof GJ, Visser P, Sierevelt IN, van Dijk CN, Kerkhoffs GM: Does perception of usefulness of arthroscopic simulators differ with levels of experience? Clin Orthop Relat Res 2011, 469:1701-1708.

[5] Michelson J: Simulation in orthopaedic education: an overview of theory and practice. J Bone Joint Surg 2006, 88:1405-1411.

[6] Satava R: Virtual reality and telepresence for military medicine. Ann Acad Med Singapore 1997, 26:118-120.

[7] Modi CS, Morris G, Mukherjee R: Computer-Simulation Training for Knee and Shoulder Arthroscopic Surgery. Arthroscopy 2010, 26:832–840.

[8] Reznek M, Harter P, Krummel T: Virtual Reality and Simulation: Training the Future Emergency Physician. Virtual real simul 2002, 9:78-87.

[9] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35:883-888.

[10] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint SurgBr 2008, 90:494-9.

[11] Brehmer M, Tolley DA: European Urology Validation of a Bench Model for Endoscopic Surgery in the Upper Urinary Tract. Eur Urol 2002, 42:175-80.

[12] Brehmer M, Swartsz R: Training on Bench Models Improves Dexterity in Ureteroscopy. Eur Urol 2005, 48:458-463.

[13] Jacomides L, Ogan K, Cadeddu JA, Pearle MS: Use of a virtual reality simulator for ureteroscopy training. J Urol 2004, 171:320-323.

[14] Knoll T, Trojan L, Haecker A, Alken P, Michel MS: Validation of Computer-Based Training in Ureterorenoscopy. BJU Int 2005, 95:1276-1279.

[15] Macmillan AI, Cuschieri A: Assessment of Innate Ability and Skills for Endoscopic Manipulations by the Advanced Dundee Endoscopic Psychomotor Tester: Predictive and Concurrent Validity. Am J Surg 1999, 177:274-277.

[16] Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD: A Novel Approach to Endourological Training: Training at the Surgical Skills Center. J Urol 2001, 166:1261-1266.

[17] Matsumoto ED, Pace KT, D'A Honey RJ: Virtual Reality Ureteroscopy Simulator as a Valid Tool for Assessing Endourological Skills. Int J Urol 2006, 13:896-901.

[18] Ogan K, Jacomides L, Shulman MJ, Roehrborn CG, Cadeddu J, Pearle MS: Virtual Ureteroscopy Predicts Ureteroscopic Proficiency of Medical Students on a Cadaver. J Urol 2004, 172:667-671.

[19] Watterson JD, Beiko DT, Kuan JK, Denstedt JD: Randomized Prospective Blinded Study Validating Acquistion of Ureteroscopy Skills Using Computer Based Virtual Reality Endourological Simulator. J Urol 2002, 168:1928-1932.

[20] Wilhelm DM, Ogan K, Roehrborn CG, Cadeddu J, Pearle MS: Assessment of Basic Endoscopic Performance Using a Virtual Reality Simulator. J Am Coll Surgeons 2002, 195:675-681.

[21] Haluck R, Satava M, Fried G, Lake C, Ritter EM, Sachdeva K, N E Seymour, Terry ML, Wilks D: Establishing a Simulation Center for Surgical Skills: What to Do and How to Do It. Surg Endosc 2007, 21:1223-232.

[22] Maura A, Galatá G, Rulli F: Establishing a Simulation Center for Surgical Skills. Surg Endosc 2008, 22(2):564.

[23] McClusky DA, Smith CD: Design and Development of a Surgical Skills Simulation Curriculum. World J Surg 2008, 32:171-181.

[24] White MA, Dehaan AP, Stephens DD, Maes A, Maatman TJ: Validation of a High Fidelity Adult Ureteroscopy and Renoscopy Simulator. J Urol 2010, 183:673-677.

[25] O'Neill PJ, Cosgarea AJ, Freedman JA, Queale WS, McFarland EG: Arthroscopic proficiency: a survey of orthopaedic sports medicine fellowship directors and orthopaedic surgery department chairs. Arthroscopy 2002, 18:795-800.

[26] Tuijthof GJ, van Stekenburg MN, Sierevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills Knee Surg Sports Traumatol Arthrosc 2010, 18:218-224.

[27] Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL: Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. J Shoulder Elbow Surg 2004, 13:196-205.

[28] Botden S, Torab F, Buzink SN, Jakimowicz JJ: The Importance of Haptic Feedback in Laparoscopic Suturing Training and the Additive Value of Virtual Reality Simulation. Surg Endosc 2008, 22:1214-1222.

[29] Kim HK, Rattner DW, Srinivasan MA: Virtual-reality-based laparoscopic surgical training: the role of simulation fidelity in haptic feedback. Comput Aided Surg 2004, 9:227-234.

[30] Moody L, Waterworth A, McCarthy AD, Harley P, Smallwood R: The feasibility of a mixed reality surgical training environment. Virtual Reality 2008, 12:77-86.

[31] Ström P, Hedman L, Särnå L, Kjellin A, Wredmark T, Felländer-Tsai L Early Exposure to Haptic Feedback Enhances Performance in Surgical Simulator Training: a Prospective Randomized Crossover Study in Surgical Residents. Surg Endosc 2006, 20:1383-1388.

[32] Korndorffer Jr JR, Dunne B, Sierra R, Stefanidis D, Touchard CL, Scott DJ: Simulator training for laparoscopic suturing using performance goals translates to the operating room. J Am Coll Surg 2005, 201:23-29.

[33] Munz Y, Almoudaris AM, Moorthy K, Dosis A, Liddle AD, Darzi AW: Curriculum-Based Solo Virtual Reality Training for Laparoscopic Intracorporeal Knot Tying: Objective Assessment of the Transfer of Skill from Virtual Reality to Reality. Am J Surg 2007, 193:774-783.

CHAPTER 6

Transfer validity of simulator training for knee arthroscopy: a randomized pilot study controlled

J.J. Stunt, G.M.M.J. Kerkhoffs, T. Pahlplatz, M. Maas, G.J.M. Tuijthof

Submitted

Abstract

Purpose: The aim of this randomized controlled pilot study is to investigate the effects of knee simulator training, to evaluate the study design, and investigate the feasibility of the used methods and procedures for later use in a large randomized controlled transfer validity study.

Methods: Thirty-one third-year medical students were recruited. Subjects that had any arthroscopic experience or received prior simulator training were excluded. Arthroscopic performance and knowledge level were compared between subjects who trained with knee simulators (investigational group) and who received theoretical training (control group). All subjects received three one-hour sessions, for three consecutive weeks. Both groups performed a knowledge quiz and an arthroscopic skill proficiency test on a cadaver knee. Outcome measures were Quiz scores, ASSET scores and Navigations scores (number of probed landmarks).

Results: As none of the participants was able to probe one landmark during the pre-test, only post-test Navigation scores could be analyzed. Participants overall improved from baseline tests with respect to knowledge level ($p=0.001$) and arthroscopic skills level ($p=0.37$). Adjusted multiple linear regression showed group effect for ASSET scores ($p=0.04$) and Navigation scores ($p=0.001$). There was no group effect for Quiz scores ($p=0.17$). Power analysis demonstrated that a sample size of 64 in each group is required for 80% power, based on a three-week training period.

Conclusion: The study provided valuable data on the trial design and outcome measures for a large-scale RCT on the effectiveness of simulator training for arthroscopy. The findings suggest that some important modifications in the study design are necessary to make a large scale randomized controlled trial feasible.

Clinical relevance: This pilot study provides information to enable a well-designed study that can provide clarity about the effectiveness and transfer validity of simulator training.

Introduction

Arthroscopy is an effective and commonly used surgical technique[1-5]. Knee arthroscopy is one of the most frequently performed orthopaedic interventional procedures[3]. Since arthroscopic procedures are complex and obtaining arthroscopic skills is challenging[6-9], adequate arthroscopic training is important. Traditionally, arthroscopic training is performed in the operating room in a one-to-one apprentice model, where the resident is allowed to perform actions under supervision[10, 11]. This is no optimal learning environment, as training time, availability of supervisors and operative experience are limited[6, 12-15], and lack of experience could compromise patient safety[14, 16-19]. Therefore, increased training efficiency and effectiveness is necessary[20, 21]. To this end, simulation training is increasingly used in surgical curricula[22-25].

An advantage of simulator training is that it facilitates supervisor-independent learning, allows the trainee to make mistakes (which potentially increases the speed of learning), and equips the resident with a basic arthroscopic skills level that would not compromise patient safety when continuing training in the operating room[9, 10, 21, 26-30]. Numerous (randomized controlled) studies have demonstrated the validity of different types of simulators[23, 31-42] and the positive effect on surgical skills[15, 43-53].

Although simulator technology has proven its benefit in many (medical) disciplines[10, 13, 54-57], scarcity of evidence on the effectiveness of arthroscopic simulators exists and most arthroscopic skills training programs do not use knee simulators routinely[58]. The potential and educational value of knee simulators has been evaluated, and high levels of internal, face, construct and content validity were demonstrated[21, 26, 31, 59-68]. The effect of knee simulator training on arthroscopic performance

was investigated as well[11, 21, 35, 69]. To our knowledge, there is only one randomized controlled study that has shown transfer validity of knee simulator training, comparing residents who trained on an arthroscopic knee simulator with a control group[70]. More convincing evidence on the benefits of knee simulators is needed to incorporate knee simulation training into arthroscopic training curricula.

The objective of this randomized controlled pilot study is to determine the following aspects: a) feasibility of the used methods and procedures to execute a full scale study, b) adequate sample size calculation, c) the appropriateness of outcome measures, and d) the effects of knee simulator training. These answers would serve a full randomized controlled transfer validity study on the use of arthroscopic simulators.

Methods

Participants

Thirty-one medical students from the Academic Medical Centre (AMC) of Amsterdam, following the course 'Musculoskeletal Disorders', voluntarily enrolled in the study in September 2015. Subjects that had any arthroscopic experience or received prior simulator training were excluded. This exclusion criterion was set to effectuate a group of participants with exactly the same baseline level and to enable a net effect measurement of simulator and theoretical training. At inclusion, demographic information of all participants was obtained. Moreover, participants reported handedness and level of videogame experience, as these aspects are known to influence simulator performance. Informed consent was obtained from all participants at study entry.

Study design

In this randomized controlled single-blinded pilot study, arthroscopic performance and knowledge level were compared between medical students who trained with knee simulators (investigational group) and who received theoretical training (control group). Participants were randomly assigned to the investigational Group 1 or the control Group 2. Transfer validity of arthroscopic skills was evaluated by a navigation task performed on a human cadaver knee joint by all participants (Figure 1A). The Arthroscopic Surgery Skill Evaluation Tool (ASSET) score[4] and the number of identified and probed landmarks were used as outcome measures. The knowledge level was evaluated by an on-line multiple choice quiz consisting of fourteen questions. Exactly the same navigation task on the human cadaveric knee joint and quiz were performed during the pre- and post-test.

Both groups performed the navigation task and the knowledge quiz to assess their baseline level (pre-test). Subsequently, all subjects received three one-hour sessions, for three consecutive weeks. We decided to distribute the training sessions (*distributed practice*) because practice interspersed with periods of rest has shown to lead to better acquisition and retention of skills than practice delivered in continuous blocks (*massed practice*)[52, 71-73]. Also, time frame and frequency of training had to be chosen such that it would fit in the study curriculum of the participants. Group 1 practiced one hour on each of three knee simulators for which face and construct validity were demonstrated[39, 40, 42, 74] (Figure 1B-D). Three different simulators were used since the study covers the effectiveness of knee simulators in general, not one specific simulator and because we were compelled to let multiple participants train simultaneously for organizational reasons. The training program consisted

of a number of tasks of varying complexity, following a proficiency-based strategy. This implied that all individuals had to perform at least at some benchmark level of performance before they could move on to the next, more complex, exercise. The simulators had similar curricula.

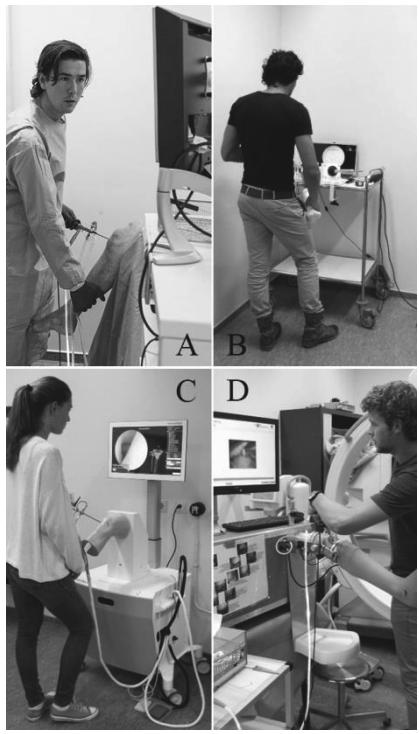


Figure 1. A: Participant performing the pre-test on a cadaver knee. B-D participants practicing on the Simendo (B), Virtamed (C) and PASSPORT (D)

VirtaMed® ArthroS (www.virtamed.com)

The VirtaMed ArthroS™ is a virtual reality simulator with tactile feedback, and provides a patient knee model and a high-end PC with touch-screen. The training program consisted of two guided navigation tasks (Guided Diagnostics I: menisci, Guided Diagnostics II:knee, two triangulation tasks (Triangulation I and Catch the stars I) and two unguided navigation taks.

Simendo simulator (www.simendo.eu)

The Simendo (Simendo, Rotterdam, The Netherlands) is a virtual reality simulator to train hand-eye coordination motor skills. The curriculum consisted of an exercise that teaches how to create space ('Creating Space'), three navigation exercises ('Four Boxes', 'Six boxes' and 'Knee navigation') one exercise for basic exploration ('Knee inspection'), and one for manipulation of the arthroscopic hook ('Knee manipulation').

PASSPORTv2[9]

The PASSPORTv2 is a combination of an anatomical model and a performance assessment system. The training consisted of four exercises: create space; a navigation task using only the scope; the same navigation task with the probe; performing a medial and lateral meniscectomy.

Group 2 was exposed to two one-hour lectures, based on literature on the anatomy of the knee, knee pathologies, description of arthroscopic techniques and equipment, pitfalls and complications[75-78]. Besides this, they attended one life knee arthroscopy. This was in all cases a meniscectomy. Finally, proficiency level and transfer of arthroscopic skills and knowledge were evaluated (post-test).

The feasibility of the used methods and procedures, sample size and outcome measures for a full scale transfer validity study were evaluated by analyzing the overall results of both pre- and post-test, and the post-test differences between the two study arms for both arthroscopic skills and knowledge.

Pre- and post-training cadaveric testing

All subjects performed a predefined navigation task on a human cadaver knee (Figure 1A). To this end, two cadaveric knees were fixated in the standard anterior arthroscopy position. A standard arthroscopy tower with a 30-degree arthroscope, arthroscopic camera, light source, and probe (Smith and Nephew, Andover Massachusetts USA) were used, as well as irrigation by means of a gravity bag. Following a brief instruction, the subjects were asked to perform the task, consisting of identifying and probing nine landmarks in the following sequence: a) medial femoral condyle, b) medial tibial plateau, c) posterior horn of medial meniscus, d) mid-section of medial meniscus, e) anterior cruciate ligament, f) lateral femoral condyle, g) lateral tibial plateau, h) posterior horn of lateral meniscus, and i) midsection of lateral meniscus[9]. Participants had five minutes to complete the navigation task, and were evaluated by an observer that was blinded to the participant's training status.

To evaluate the navigation task, two pre-test and two post-test observers filled out an Arthroscopic Surgery Skill Evaluation Tool (ASSET)[4]. The ASSET is a rating scale for the assessment of global arthroscopic technical skill, applicable to multiple arthroscopic procedures in both the live OR and simulated environments. It includes eight domains with a maximum of 38 points in total, considering safety, view, dexterity, flow and quality of procedure and autonomy, which can be evaluated by a 5-point Likert-type scale with descriptors at 1, 3, and 5[4]. These numbers are not arbitrary, but correspond to the levels of the Dreyfus model of skill acquisition with "1" representing the novice level, "3" representing the competent level, and "5" representing the expert level of arthroscopic skill performance[79, 80]. The ASSET has demonstrated to be a useful, valid, and reliable method for assessing surgeon performance of diagnostic knee

arthroscopy in cadaveric specimens[4]. Each subject was assessed by one observer. Besides the ASSET, for each participant the number of landmarks in the knee identified and probed within five minutes was scored.

Pre- and post-training knowledge testing

Participants performed a pre- and post-training on-line quiz. The quiz examined the level of knowledge about anatomical structures, pathologies, handling of instruments and placements of portals. The quiz score had a range from zero to a hundred points.

Statistical analysis

Data were analyzed with the Statistical Package for the Social Sciences version 22.0 (SPPS, Chicago, IL) and R. As one knee was qualitatively better -and thus easier to perform a navigation task on- than the other, this was accounted for in the statistical analysis (referred to as 'low' or 'high' knee complexity). Level of videogame experience was categorized as either none, moderate or extensive. The number of landmarks identified and probed was referred to as 'Navigation score'.

Normality of the parameters was tested using the Shapiro–Wilk test. Means and standard deviations were calculated for parametrical continuous data (age, Quiz scores), medians and ranges were calculated for non-parametrical continuous data (number of structures and ASSET scores) whereas frequencies and proportions were calculated for categorical and data (handedness, sex, computer game experience and knee complexity). Baseline characteristics and potential chance imbalances between Group 1 and Group 2 were determined using the

Student's t-test and Mann-Whitney U test for continuous variables and the chi-squared test categorical variables.

The paired samples t-test was performed to analyze differences between pre-and post-test Quiz scores. The Wilcoxon signed rank test was performed to check for differences between pre- and post-test ASSET scores. Linear regression analyses were used to determine differences between the randomized groups at the post-test for outcome measures in four stages: (1) raw group effects on outcome measures; (2) adjustment for individual covariates (sex, handedness, computer game experience and knee complexity); (3) adjustment for knee complexity; and (4) interaction effects model to examine group by time (baseline scores and post-test scores) effect. To meet assumptions for linear regression, ASSET and Navigation scores were log-transformed. The regression coefficient (B) and the explained variance R^2 (determined by squaring the correlation coefficient R) were calculated. Correlation coefficients ≤ 0.35 were considered to represent weak correlations, 0.36 to 0.67 moderate correlations, and 0.68 to 1.0 high correlations, with coefficients ≥ 0.91 very high correlations[81].

P-values of 0.05 were considered statistically significant. In case of non-significant results, a power calculation was performed, in order to investigate the number of participants required for a definitive RCT to detect a significant effect on outcome measures. The power calculation is based on a) median ASSET scores and b) the clinical relevant difference in ASSET score.

Results

Thirty-one students were recruited. Six participants failed to complete the entire study procedure (Figure 2). Table 1 presents the characteristics of

the subjects included in the analysis. Statistical differences were not observed between the two groups at baseline with respect to demographic characteristics and videogame experience. During the pre-test, more Group 1 than Group 2 participants performed the cadaver test on the 'low'-complexity knee (8:5). During the post-test, more Group 1 than Group 2 participants performed the cadaver test on the 'high'-complexity knee (7:4).

Wilcoxon signed rank test showed participants overall improved from baseline tests with respect to knowledge and arthroscopic skills level (Table 2). As none of the participants was able to identify and probe one structure within five minutes during the pre-test (and thus, no quantity could be measured), only post-test Navigation scores could be analyzed. Overall median score for number of structures was 6. Linear regression analyses were performed to investigate the association between group number and outcome measures, and whether one or more individual covariates could have influenced that association, i.e. if there was confounding or effect-modification.

Univariate linear regression analyses showed that there was a raw group effect on ASSET, Navigation and Quiz score, but the association was weak and non-significant, as the explained variance (R^2=0.001, 0.003 and 0.07 respectively) and the p-value (p=0.88, 0.78 and 0.17 respectively) demonstrated. Coefficient estimates (B) showed that post-test ASSET score decreased with nearly one scale point in the control group (B=-0.98), the number of structures was almost one more for the control group (B=0.93). Post-test Quiz score showed a decrease of 6.4 points in the control group (B=-6.4). Further univariate linear regression analysis revealed that outcome measures were not dependent on individual covariates, except for knee complexity (R^2=0.19, p=0.03)

(Table 3). Median ASSET and Navigation scores were significantly higher for participants that performed the post-test on the low complexity knee (4.5 ASSET points more, p=0.03 and four structures more on Navigation score, p=0.01 respectively) (Table 3). Because of the significant influence of knee complexity on arthroscopic performance, it was included in the multiple regression model.

Multiple linear regression showed that adjustment for knee complexity, adjustment for baseline scores and/or adjustment for both baseline and score knee complexity inflated the association (correlation coefficient) between group number and outcome measures (Table 4): although correlations between group number and ASSET scores remained weak to moderate, they increased compared to the model for group effect alone (stage 1) (p=0.04). Also the association between group effect adjusted for knee complexity and Navigation scores increased (p=0.01). The association between group number and Quiz scores increased when adjusted for baseline scores, but remained weak and non-significant.

As most results were non-significant, a sample size calculation was performed. Based on the median ASSET scores and the clinical relevant difference in ASSET scores, a sample size of 64 participants in each group will have 80% power to detect a significant difference in ASSET score if we preserve a study period of three weeks. A number of 80 participants per group will increase the power and will overcome the problem of possible study drop-outs.

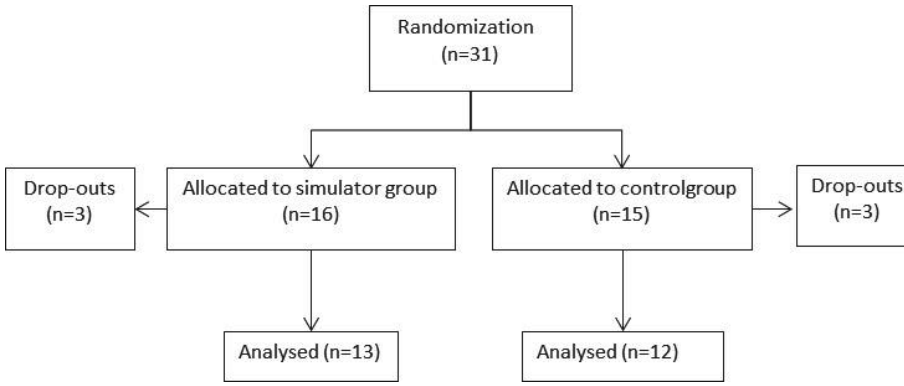


Figure 2: Flow diagram outlining enrollment, allocation and assessments of participants

Table 1: Demographic data for the simulator and control group			
Parameter	**Simulator group**	**Control group**	**p-value**
Sex (male: female)	7:6	7:5	0.83
Age (mean (sd))	21.5 (1.6)	21.8 (1.5)	0.65
Handedness (right: left)	11:2	10:2	0.93
Videogame experience	10:0:3	5:2:5	0.13

Table 2: pre- and post-test scores			
Score	**Pre-test**	**Post-test**	**p-value**
Quiz mean (sd)	34 (11.8)	49 (13.5)	<0.001
Navigation score	-	6 (1-18)	-

Table 3: association between individual covariates and outcome measures										
		ASSET*			Navigation*			Quiz		
		B (95%CI)	**R^2**	**p-value**	**B (95%CI)**	**R^2**	**p-value**	**B (95%CI)**	**R^2**	**p-value**
Group		-0.98(-1.26-0.76)	0.001	0.88	0.93 (-1.63-0.53)	0.003	0.78	-6.4 (-15.8-3.0)	0.07	0.17
Sex		-0.89(-1.12-0.69)	0.05	0.29	-0.75 (-1.29-0.43)	0.05	0.29	0.004	0.004	0.75
Handedness		-0.80(-1.1-0.57)	0.08	0.17	-0.73 (-1.57-0.34)	0.03	0.2	-	-	-
Computer games	**moderate**	0.90(-0.55-1.41)	0.05	0.61	0.51 (-0.18-1.47)	0.08	0.38	-	-	-
	extensive	0.88(-0.67-1.16)			1.08 (-0.59-1.99)					
Knee complexity		-0.78(-0.98--0.62)	0.24	0.03	-0.48(-0.77--0.29)	0.31	0.004	-	-	-

*for interpretation of the regression coefficients, ASSET and navigation task scores were back transformed

Table 4: association between group number and outcome measures, adjusted for knee complexity and baseline score

	ASSET*			Navigation*			Quiz		
	B (95%CI)	**R^2**	**p-value**	**B (95%CI)**	**R^2**	**p-value**	**B (95%CI)**	**R^2**	**p-value**
Group + knee complexity	-0.93 (-1.17-0.740)	0.20	0.08	-0.79 (-1.28-0.47)	0.34	0.01	-	-	-
Group + baseline score							-6.2(-15.9-3.3)	0.0.7	0.397
Group + baseline score + knee complexity	-0.98 (-1.23-0.78)	0.32	0.04				-	-	-

*for interpretation of the regression coefficients, ASSET and Navigation scores were back transformed

Discussion

The purpose of conducting this study was to carefully investigate the study design, applied methods, outcome measures and preliminary results on the effect of knee simulator training, before moving to a large-scale randomized trial. Statistical analysis has demonstrated overall progression between pre- and post-test scores, with significant progression for Quiz scores. By lack of pre-test Navigation scores, quantitative comparison between pre- and post-test results was not possible. However, it is legitimate to state that there was significant progress in both groups as the median post-test Navigation score was 6 points. Group number on its own had no significant effect on outcomes indicating no effect between simulator versus no simulator. But the linear regression analysis demonstrated that association between group number and outcome measures increased substantially when baseline scores and knee complexity were included as independent variables in the regression model. The results of this study suggest a positive effect of simulator training on arthroscopic skills- and knowledge level. However, judging from these results, some methodological modifications are necessary in order to create an optimal trial design for a future large-scale RCT.

Firstly, skills progression on the ASSET scale was not significant. The rationale for using the ASSET was to have an expert opinion on participant's performance, besides a quantitative measure. However, the ASSET appeared not sufficiently sensitive to measure a significant progression after a three-week training period. Moreover, we noted a floor effect, as a large part of the participants got the lowest possible score and none of the participants reached a competent level. As a consequence, participants could not be distinguished from each other sufficiently, which limited the reliability of the results. For a future RCT, a possible

alternative is the application of a tool to assess learning curves. A number of studies describe such tools and indicate that skills progression can be measured in a range of five to 30 cases[82-85]. Van Oldenrijk and co-workers used time-action analysis to assess a learning curve, and demonstrated learning took place within the first five to ten procedures[85]. As a more sensitive alternative to the ASSET, we propose to establish learning curves by using time-action analysis, measuring the number of actions and duration of each action during the navigation task of the diagnostic phase of a knee arthroscopy. Based on the minimum amount of cases to establish a learning curve and weekly training sessions, we advocate a training period of at least five weeks.

Secondly, our results indicate that the use of the Navigation score as an outcome measure should be reconsidered. We had to abandon the original idea of using task time as an outcome measure, as none of the participants was able to complete one navigation round within five minutes during the pre-test and only 40% of the participants managed to complete one sequence during the post-test (Table 2). Thus, this task might not be very appropriate for the assessment of novices. However, task time has previously shown to be a valid, and easy to understand and apply outcome measure to assess surgical skills learning, and is a commonly used and validated criterion to measure the validity and effectiveness of simulator training[9, 29, 39, 40, 68, 73, 74, 86-88]. Therefore, we would recommend task time as outcome measure in a full-scale RCT, but in a modified manner. Options are to extend the time limit to ten minutes, to assess a guided navigation task or to place the arthroscope and probe in the lateral compartment by a supervisor before starting the task. However, if participants have no or very little knowledge of the anatomy of the knee, these options are not applicable, and the alternative is to use task

time only as a post-test outcome measure to investigate differences between the simulator and the control group.

Lastly, the study was not sufficiently powered to detect small effect sizes. We expect the differences in arthroscopic performance before and after simulator training will be statistically significant with increased power. Based on the current study design, a number of at least 64 in each study arm is required to find significant results. However, using a more sensitive assessment tool and including more training sessions, a number of 25 participants per group is justified. This was also shown by Cannon and co-workers[70].

The present study has some limitations. Firstly, four different supervisors filled out the ASSET score, and participants were assessed by a single one of them. Ideally, multiple raters, trained with distinct examples to increase consensus, should assess both pre-test and post-tests scoring of each subject[4]. Unfortunately, this was not feasible for the current study, due to limited availability of eligible assessors. Secondly, a relatively large part of the simulator group compared to the control group was randomly assigned to perform the cadaver test on the 'low'-complexity knee during the pre-test and on the 'high'-complexity knee during the post-test. This coincidence significantly distorted the results, as ASSET and Navigation scores on the 'high'-complexity knee were significantly lower than on the 'low'-complexity knee, and partly explains the lack of differences in Navigation score between Group 1 and 2. Since knee quality influences arthroscopic performance, cadaver knees of the same quality should ideally be used in a future study, or it should be ensured that knees of different quality are evenly distributed among the two groups.

Some trends were observed in this study. With the above mentioned modifications, a more reliable basis for a full-scale RCT on simulator training is provided. An adequately powered RCT with a prolonged training period and better aligned outcome measures is highly recommended to rightly inform orthopedic trainers about the effectiveness and transfer validity of knee simulators, the implementation of a simulation curriculum in residency programs and whether simulator training can contribute to prevent residents in training from compromising patient safety in the operating room.

Conclusion

The study provided valuable data on the trial design and outcome measures for a large-scale RCT on the effectiveness of simulator training for arthroscopy. The findings suggest that important modifications in the study design are necessary to make a large scale randomized controlled trial feasible. Before moving to a definitive RCT it is imperative that duration of training, sample size, outcome measures and assessment of proficiency level are optimized. A study period of minimally five weeks, with 25 participants included in each study arm, using time-action analysis to establish learning curves, an adjusted navigation task and multiple, trained, raters, assessing both pre-test and post-tests scoring of each subject, should enable a study that can provide clarity about the effectiveness and transfer validity of simulator training.

References

[1] Garrett Jr. WE, Swiontkowski MF, Weinstein JN, Callaghan J, Rosier RN, Berry DJ, Harrast J, Derosa GP: American Board of Orthopaedic Surgery Practice of the Orthopaedic Surgeon: Part-II, certification examination case mix. J Bone Joint Surg Am 2006, 88:660-667.

[2] Hobby J, Griffin D, Dunbar M, Boileau P: Is arthroscopic surgery for stabilisation of chronic shoulder instability as effective as open surgery? A systematic review and meta-analysis of 62 studies including 3044 arthroscopic operations. J Bone Joint SurgBr 2007, 89:1188-1196.

[3] Kim S BJ, Meehan JP, Jamali A, Marder R: Increase in outpatient knee arthroscopy in the United States: a comparison of National Surveys of Ambulatory Surgery, 1996 and 2006. J Bone Joint SurgAm 2011, 93:994-1000.

[4] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner CD, Garrett WE, Olson T, Warme WJ, Nicandri GT: The Arthroscopic Surgical Skill Evaluation Tool (ASSET). Am J Sports Med 2013, 41:1229-1237.

[5] Zhang YZ: Innovations in Orthopedics and Traumatology in China. Chin MedJ(Engl) 2015, 128:2841-2844.

[6] Hodgins JL, Veillette C, Biau D, Sonnadara R: The knee arthroscopy learning curve: quantitative assessment of surgical skills. Arthroscopy 2014, 30:613-621.

[7] Insel A, Carafino B, Leger R, Arciero R, Mazzocca AD: The development of an objective model to assess arthroscopic performance. J Bone Joint Surg Am 2009, 91:2287-2295.

[8] O'Neill PJ, Cosgarea AJ, Freedman JA, Queale WS, McFarland EG: Arthroscopic proficiency: a survey of orthopaedic sports medicine fellowship directors and orthopaedic surgery department chairs. Arthroscopy 2002, 18:795-800.

[9] Tuijthof GJ, van Sterkenburg MN, Sierevelt IN, van OJ, van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills. KneeSurgSports TraumatolArthrosc 2010, 18:218-224.

[10] Michelson JD: Simulation in orthopaedic education: an overview of theory and practice. J Bone Joint Surg Am 2006, 88:1405-1411.

[11] Rebolledo BJ, Hammann-Scala J, Leali A, Ranawat AS: Arthroscopy skills development with a surgical simulator: a comparative study in orthopaedic surgery residents. Am J Sports Med 2015, 43:1526-1529.

[12] Kairys JC, McGuire K, Crawford AG, Yeo CJ: Cumulative operative experience is decreasing during general surgery residency: a worrisome trend for surgical trainees? J Am Coll Surg 2008, 206:804-811.

[13] Kunkler K: The role of medical simulation: an overview. IntJMedRobot 2006, 2:203-210.

[14] Peabody T, Nestler S, Marx C, Pellegrini V: Resident duty-hour restrictions-who are we protecting? AOA critical issues. J Bone Joint SurgAm 2012, 94:e131.

[15] Sonnadara RR, Van Vliet A, Safir O, Alman B, Ferguson P, Kraemer W, Reznick R: Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. Surgery 2011, 149:745-749.

[16] Hall MP, Kaplan KM, Gorczynski CT, Zuckerman JD, Rosen JE: Assessment of arthroscopic training in U.S. orthopedic surgery residency programs--a resident self-assessment. Bull NYU Hosp Jt Dis 2010, 68:5-10.

[17] Peabody T: The effect of work hour restrictions on the education of orthopaedic surgery residents. Clin Orthop Relat Res 2006, 449:128-133.

[18] Simpson C, Cottam H, Fitzgerald JE, Giddings CE: The European working time directive has a negative impact on surgical training in the UK. Surgeon 2011, 9:56-57.

[19] Zuckerman JD, Kubiak EN, Immerman I, Dicesare P: The early effects of code 405 work rules on attitudes of orthopaedic residents and attending surgeons. J Bone Joint Surg Am 2005, 87:903-908.

[20] Bridges M, Diamond DL: The financial impact of teaching surgical residents in the operating room. Am J Surg 1999, 177:28-32.

[21] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Br 2008, 90:494-499.

[22] Atesok K, Mabrey JD, Jazrawi LM, Egol KA: Surgical simulation in orthopaedic skills training. JAmAcadOrthopSurg 2012, 20:410-422.

[23] Braman JP, Sweet RM, Hananel DM, Ludewig PM, Van Heest AE: Development and validation of a basic arthroscopy skills simulator. Arthroscopy 2015, 31:104-112.

[24] Karam MD, Pedowitz RA, Natividad H, Murray J, Marsh JL: Current and future use of surgical skills training laboratories in orthopaedic resident education: a national survey. J Bone Joint Surg Am 2013, 95:e4.

[25] Sanfey HA, Dunnington GL: Basic surgical skills testing for junior residents: current views of general surgery program directors. J Am Coll Surg 2011, 212:406-412.

[26] Cannon WD ED, Garrett Jr. WE, Hunter RE, Sweeney HJ: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res 2006, 442:21-29.

[27] Heng PA, Cheng CY, Wong TT, Wu W, Xu Y, Xie Y, Chui YP, Chan KM, Leung KS: Virtual reality techniques. Application to anatomic visualization and orthopaedics training. Clin Orthop Relat Res 2006, 442:5-12.

[28] Mabrey JD, Cannon WD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Mevis H, Garrett WE, Poss R: Development of a virtual reality arthroscopic knee simulator. Stud Health Technol Inform 2000, 70:192-194.

[29] McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR: Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res 2006, 442:13-20.

[30] Safir O, Dubrowski A, Mirsky L, Lin C, Backstein D, Carnahan A: What skills should simulation training in arthroscopy teach residents? International journal of computer assisted radiology and Surgery 2008:433-437.

[31] Fucentese SF, Rahm S, Wieser K, Spillmann J, Harders M, Koch PP: Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. Knee Surg Sports Traumatol Arthrosc 2015, 23:1077-1085.

[32] Martin KD, Belmont PJ, Schoenfeld AJ, Todd M, Cameron KL, Owens BD: Arthroscopic basic task performance in shoulder simulator model correlates with similar task performance in cadavers. J Bone Joint SurgAm 2011, 93:e1271-e1275.

[33] Martin KD, Cameron K, Belmont PJ, Schoenfeld A, Owens BD: Shoulder arthroscopy simulator performance correlates with resident and shoulder arthroscopy experience. JBone Joint SurgAm 2012, 94:e160.

[34] Martin KD, Patterson D, Phisitkul P, Cameron KL, Femino J, Amendola A: Ankle Arthroscopy Simulation Improves Basic Skills, Anatomic Recognition, and Proficiency During Diagnostic Examination of Residents inTraining. Foot and Ankel International 2015, 36:827-835.

[35] Olson T, Koehler R, Butler A, Amsdell S, Nicandri G: Is there a valid and reliable assessment of diagnostic knee arthroscopy skill? Clin Orthop Relat Res 2013, 471:1670-1676.

[36] Palter VN, Grantcharov TP: Development and validation of a comprehensive curriculum to teach an advanced minimally invasive procedure: a randomized controlled trial. Ann Surg 2012, 256:25-32.

[37] Rose K, Pedowitz R: Fundamental arthroscopic skill differentiation with virtual reality simulation. Arthroscopy 2015, 31:299-305.

[38] Stegemann AP, Ahmed K, Syed JR, Rehman S, Ghani K, Autorino R, Sharif M, Rao A, Shi Y, Wilding GE, Hassett JM, Chowriappa A, Kesavadas T, Peabody JO, Menon M, Kaouk J, Guru KA: Fundamental skills of robotic surgery: a multi-institutional randomized controlled trial for validation of a simulation-based curriculum. Urology 2013, 81:767-774.

[39] Stunt JJ, Kerkhoffs GM, Horeman T, van Dijk CN, Tuijthof GJ: Validation of the PASSPORT V2 training environment for arthroscopic skills. Knee Surg Sports Traumatol Arthrosc 2014, 24: 2038-2045.

[40] Stunt JJ, Kerkhoffs GM, van Dijk CN, Tuijthof GJ: Validation of the ArthroS virtual reality simulator for arthroscopic skills. KneeSurgSports TraumatolArthrosc 2015, 23:3436-3442.

[41] Valdis M, Chu MW, Schlachta CM, Kiaii B: Validation of a Novel Virtual Reality Training Curriculum for Robotic Cardiac Surgery: A Randomized Trial. Innovations(Phila) 2015, 10:383-388.

[42] Verdaasdonk EG, Stassen LP, Monteny LJ, Dankelman J: Validation of a new basic virtual reality simulator for training of basic endoscopic skills: the SIMENDO. Surg Endosc 2006, 20:511-518.

[43] Aim F, Lonjon G, Hannouche D, Nizard R: Effectiveness of Virtual Reality Training in Orthopaedic Surgery. Arthroscopy 2015, 32:224-232.

[44] Franzeck FM, Rosenthal R, Muller MK, Nocito A, Wittich F, Maurus C, Dindo D, Clavien PA, Hahnloser D: Prospective randomized controlled trial of simulator-based versus traditional in-surgery laparoscopic camera navigation training. Surg Endosc 2012, 26:235-241.

[45] Grantcharov TP, Kristiansen VB, Bendix J, Bardram L, Rosenberg J, Funch-Jensen P: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. BrJ Surg 2004, 91:146-150.

[46] Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR: Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. Br J Surg 2008, 95:1088-1097.

[47] Henn RF 3rd, Shah N, Warner JJ, Gomoll AH: Shoulder arthroscopy simulator training improves shoulder arthroscopy performance in a cadaveric model. Arthroscopy 2013, 29:982-985.

[48] Jackson WF, Khan T, Alvand A, Al-Ali S, Gill HS, Price AJ, Rees JL: Learning and retaining simulated arthroscopic meniscal repair skills. The Journal of Bone and Joint Surgery: American Volume 2012, 94:1-8.

[49] Palter VN: Comprehensive training curricula for minimally invasive surgery. JGradMedEduc 2011, 3:293-298.

[50] Palter VN, Grantcharov T, Harvey A, Macrae HM: Ex vivo technical skills training transfers to the operating room and enhances cognitive learning: a randomized controlled trial. Ann Surg 2011, 253:886-889.

[51] Sedlack RE, Kolars JC: Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 2004, 99:33-37.

[52] Sonnadara RR, Garbedian S, Safir O, Nousiainen M, Alman B, Ferguson P, Kraemer W, Reznick R: Orthopaedic Boot Camp II: examining the retention rates of an intensive surgical skills course. Surgery 2012, 151:803-807.

[53] Zendejas B, Cook DA, Bingener J, Huebner M, Dunn WF, Sarr MG, Farley DR: Simulation-based mastery learning improves patient outcomes in laparoscopic inguinal hernia repair: a randomized controlled trial. Ann Surg 2011, 254:502-509.

[54] Block EF, Lottenberg L, Flint L, Jakobsen J, Liebnitzky D: Use of a human patient simulator for the advanced trauma life support course. Am Surg 2002, 68:648-651.

[55] Byrne AJ, Jones JG: Inaccurate reporting of simulated critical anaesthetic incidents. Br J Anaesth 1997, 78:637-641.

[56] Kohn LT, Corrigan JM, Donaldson MS: To err is human:building a safer health system. Washington, DC.: National Academy Press, 1999.

[57] Lynagh M, Burton R, Sanson-Fisher R: A systematic review of medical skills laboratory training: where to from here? Med Educ 2007, 41:879-887.

[58] Akhtar KS, Chen A, Standfield NJ, Gupte CM: The role of simulation in developing surgical skills. Curr Rev Musculoskelet Med 2014, 7:155-160.

[59] Bliss JP, Hanner-Bailey HS, Scerbo MW: Determining the efficacy of an immersive trainer for arthroscopy skills. Stud Health Technol Inform 2005, 111:54-56.

[60] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35:883-888.

[61] Gomoll AH, Pappas G, Forsythe B, Warner JJ: Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. Am J Sports Med 2008, 36:1139-1142.

[62] McCarthy JC, Jibodh SR, Lee JA: The role of arthroscopy in evaluation of painful hip arthroplasty. ClinOrthopRelat Res 2009, 467:174-180.

[63] Modi CS, Morris G, Mukherjee R: Computer-simulation training for knee and shoulder arthroscopic surgery. Arthroscopy 2010, 26:832-840.

[64] Pedowitz RA, Esch J, Snyder S: Evaluation of a virtual reality simulator for arthroscopy skills development. Arthroscopy 2002, 18:E29.

[65] Schmidt RA: Motor schema theory after 27 years: reflections and implications for a new theory. Res Q Exerc Sport 2003, 74:366-375.

[66] Smith S, Wan A, Taffinder N, Read S, Emery R, Darzi A: Early experience and validation work with Procedicus VA--the Prosolvia virtual reality shoulder arthroscopy trainer. Stud Health Technol Inform 1999, 62:337-343.

[67] Tashiro Y, Miura H, Nakanishi Y, Okazaki K, Iwamoto Y: Evaluation of skills in arthroscopic training based on trajectory and force data. Clin Orthop Relat Res 2009, 467:546-552.

[68] Tuijthof GJ, Visser P, Sierevelt IN, van Dijk CN, Kerkhoffs GM: Does perception of usefulness of arthroscopic simulators differ with levels of experience? Clin Orthop Relat Res 2011, 469:1701-1708.

[69] Jacobsen ME, Andersen MJ, Hansen CO, Konge L: Testing basic competency in knee arthroscopy using a virtual reality simulator: exploring validity and reliability. J Bone Joint Surg Am 2015, 97:775-781.

[70] Cannon WD, Nicandri GT, Reinig K, Mevis H, Wittstein J: Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. J Bone Joint Surg Am 2014, 96:e57.

[71] Donovan JJ, Radosevich DJ: A meta-analytic review of the distribution of practice effect: now you see it, now you don't. J of Appl Psych 1999, 84:795-805.

[72] Lee TD, Genovese ED: Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. Research Quarterly for Exercise and Sports 1988, 60:59-65.

[73] Moulton CA, Dubrowski A, Macrae H, Graham B, Grober E, Reznick R: Teaching surgical skills: what kind of practice makes perfect?: a randomized, controlled trial. AnnSurg 2006, 244:400-409.

[74] Verdaasdonk EG, Stassen LP, Schijven MP, Dankelman J: Construct validity and assessment of the learning curve for the SIMENDO endoscopic simulator. Surg Endosc 2007, 21:1406-1412.

[75] Allum R: Complications of arthroscopy of the knee. J Bone Joint Surg Br 2002, 84:937-45.

[76] Caspari RB: Arthroscopy--an orthopaedic surgical subspecialty or a technique? Arthroscopy 1991, 7:390-393.

[77] Nederlandse Vereniging voor Arthroscopie: Syllabus Scopiecursus 2: Basiscursus arthroscopie van de knie. 2000.

[78] Verhaar JAN, van Mourik JB: Leerboek Orthopedie: Bohn Stafleu van Loghum, 2008.

[79] Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S: General competencies and accreditation in graduate medical education. Health Aff(Millwood) 2002, 21:103-111.

[80] Carraccio CL, Benson BJ, Nixon LJ, Derstine PL: From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. Acad Med 2008, 83:761-767.

[81] Taylor R: Interpretation of the Correlation Coefficient: A Basic Review. Journal of Diagnostic Medical Sonography 1990, 6:35-39.

[82] Castricini R, De Benedetto M, Orlando N, Rocchi M, Zini R, Pirani P: Arthroscopic Latarjet procedure: analysis of the learning curve. Musculoskelet Surg 2013, 97 Suppl 1:93-98.

[83] Guttmann D, Graham RD, MacLennan MJ, Lubowitz JH: Arthroscopic rotator cuff repair: the learning curve. Arthroscopy 2005, 21:394-400.

[84] Hoppe DJ, de Sa D, Simunovic N, Bhandari M, Safran MR, Larson CM, Ayeni OR: The learning curve for hip arthroscopy: a systematic review. Arthroscopy 2014, 30:389-397.

[85] van Oldenrijk J, Schafroth MU, Bhandari M, Runne WC, Poolman RW: Time-action analysis (TAA) of the surgical technique implanting the collum femoris preserving (CFP) hip arthroplasty. TAASTIC trial identifying pitfalls during the learning curve of surgeons participating in a subsequent randomized controlled trial (an observational study). BMC Musculoskelet Disord 2008, 9:93.

[86] Monahan E, Shimada K: Verifying the effectiveness of a computer-aided navigation system for arthroscopic hip surgery. Stud Health TechnolInform 2008, 132:302-307.

[87] Newmark J, Dandolu V, Milner R, Grewal H, Harbison S, Hernandez E: Correlating virtual reality and box trainer tasks in the assessment of laparoscopic surgical skills. Am J Obstet Gynecol 2007, 197(5):546.e1-4

[88] Panait L, Bell RL, Roberts KE, Duffy AJ: Designing and validating a customized virtual reality-based laparoscopic skills curriculum. J Surg Educ 2008, 65:413-417.

CHAPTER 7

Variation in joint stressing magnitudes during knee arthroscopy

J.J. Stunt, P.H.L.M. Wulms, G.M.M.J. Kerkhoffs, I.N. Sierevelt, M.U. Schafroth, G.J.M. Tuijthof

Knee Surg Sports Traumatol Arthrosc (2014) 22:1529–153

Abstract

Purpose: When performing knee arthroscopy, joint stressing is essential to increase the operative joint space. Adequate training of joint stressing is important, since high stressing forces can damage knee ligaments and low stressing might not give sufficient operative space. As forces are difficult to transfer since they cannot be seen, simulators might be suited to train joint stressing as they can visualize the amount of applied stress. This requires the joint stressing thresholds to be validated. The purpose of this study was to measure the variation of the maximum joint stressing forces applied by various surgeons in vivo in a human population, and based on that derive thresholds for safe stressing.
Methods: From studies on ligament failure properties we inferred a theoretical maximum stressing force of 78 N. Twenty-one patients were included and knee arthroscopies were performed by five experienced surgeons. Forces solely performed in the varus and in valgus direction were measured. A load sensor was mounted on a belt, which was rotated along the hip to measure both varus and valgus stressing. The measurements started as soon as the interior of the knee joint was visualized using joint stressing.
Results: The average maximum stressing force was 60 N (SD= 28 N). The mean first frame force was 47 N (SD=34 N). No significant differences were found between varus and valgus stressing.
Conclusion: Since variation in stressing forces is high, offering training cases on simulators where the complete range of stressing forces can be experienced is recommended. Abiding to safety levels is essential to increase patient safety.

This research was funded by the Marti-Keuning Eckhart Foundation, Lunteren, The Netherlands.

Introduction

During knee arthroscopies, the lower leg is stressed to increase the available joint space. Stressing the lower leg in valgus direction increases the space in the medial compartment and stressing in varus direction increases the space in the lateral compartment. This way, the surgeon has increased space for inspection, navigation and therapeutic treatment in the complete knee joint. The simultaneous performance of joint stressing and triangulation is a complex task, especially as the surgeon watches the arthroscopic view on the monitor during this process to verify his actions and judge proper stressing[1]. Thus, stressing and triangulation are predominantly executed based on haptic sensory feedback.

Learning to apply the load adequately is challenging, as the stressing forces applied cannot be seen; it is difficult for the resident to judge what force to apply, and for the surgeon to transfer his/her skills. This complicates skills training and potentially compromises training effectiveness and patient's safety. Safe execution of joint stressing is important, since too high stressing loads can possibly damage the collateral and/or cruciate ligaments[2-4]. Simulators can facilitate this specific type of skills training, since they provide built-in sensors that can measure performance, and thus the applied stressing forces. They can objectively judge efficiency or quality of a performance by measuring the path length and movement of the arthroscope and probe, or by counting the amount of errors made[5, 6]. This gives the benefits that residents can practice joint stressing by actual handling without compromising patient safety[5, 7-11], and that feedback on stressing forces can be offered by visual or audible cues[6, 11-13]. To accomplish this, joint stressing thresholds need to be validated.

To the best of our knowledge, no safe loading levels for joint stressing have been proposed, but they are needed to offer validated simulator training. A study of Schmid et al.[4] reported maximum stressing forces during knee arthroscopy, but their data cannot be used as seven patients in that study had a torn ligament; this study aims at investigating joint stressing forces in patient with normal intact ligaments.

Thus, the aim of this study is to measure quantitative magnitudes and the variation of joint stressing forces during arthroscopic knee surgery performed by experienced surgeons. It is important to measure these stressing forces during arthroscopy, because A) the patients are under general or spinal anaesthesia and will not limit the maximum stressing load by their pain limit and B) the orthopaedic surgeon determines the stressing magnitude based on visual information received from the arthroscopic view. A theoretical safe joint stressing level will be deduced. The results will be used to derive thresholds for safe stressing and recommendations for effective stressing training.

Materials and Methods

Theoretical maximum allowable stressing force (TMSF)

To determine the maximum allowable joint stressing force (TMSF), the longest expected leg length (Figure 1, L1), the length of the medial collateral ligament (MCL, Figure 1, L2) and the maximal allowable force on the weakest set of collateral ligaments need to be determined (F2). Using these measures, the joint stressing maximum performed is derived at the ankle level (Figure 1, F1), which is the location that is routinely used to apply stressing forces by the surgeon. When stressing is performed at magnitudes lower than the TMSF, no damage of the knee ligaments is expected in 95% of the adult population.

Several studies on cadavers determined the failure properties of the MCL[14-16] (Table 1). The subtract of 2 times the standard deviation from the measured mean values, was defined as the strength of the weakest MCL for each of the datasets of three papers (Table 1, Column 3). All values found in literature were pooled by calculating the mean value of these weakest tensile strengths resulting in 427 N. This force was multiplied by a factor 2, because during knee stressing also the other ligaments such as the cruciate ligaments and joint capsules bear part of the load. Thus, a tensile force of 854N was set as the load that causes the weakest MCL at the 95% percentile of the population to fail.

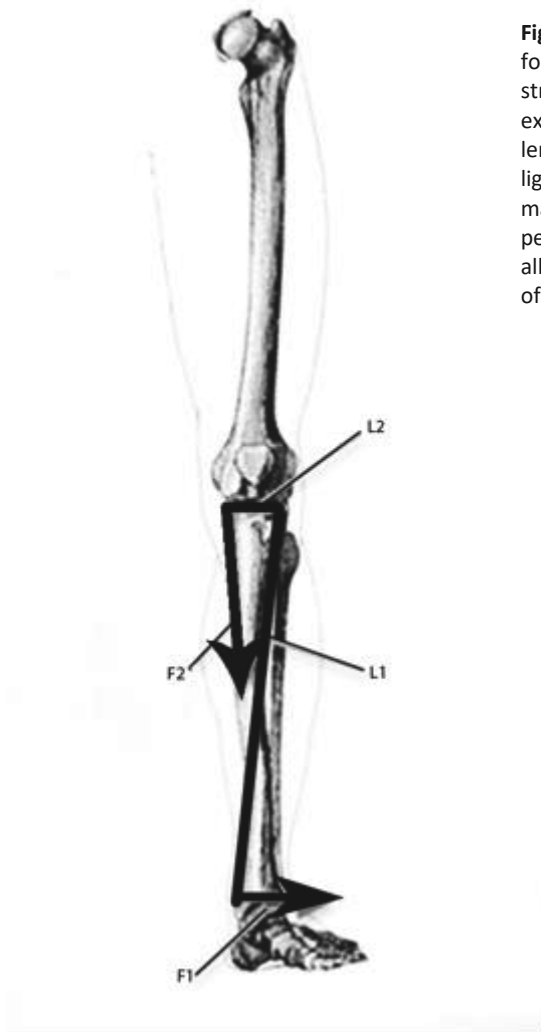


Figure 1 Moment arms of and forces on the lower leg during join stressing. L1 is the longest expected leg length; L2 is the length of the medial collateral ligament; F1 is the level where maximum joint stressing is performed; F2 is the maximum allowable force on the weakest set of collateral ligaments

Subsequently, the TMSF at the ankle joint level was calculated by multiplying this tensile force with the ratio of the contact point and the condyle and the leg length. The ratio affects the magnitude of TMSF. Therefore, the average leg length for both men and women were found, as well as the mean width of femoral condyles, the mean moment arm of the

MCL and the ratio of the contact point and the condyle (Table 2). No differences were found between the ratios of leg length and the width of the condyle. The length of the moment arm of the MCL was calculated by means of the ratio between the distance from the pivot point and the attachment of the MCL, and the width of the condyle of the femur, which were obtained from postero-anterior radiographs from the knee. This ratio was 3:4[17]. The ratio between the force on the lower leg and the force on the MCL is calculated by division of the lower leg length and the moment arm of the MCL (Table 2: last row), which was 11. Thus, the force on the lower leg applied at the ankle level causes a force on the MCL that is 11 times larger. To calculate the accompanying joint stressing force at ankle level, 854N was divided by 11 (Table 2D), which gave the theoretical allowable stressing force (TMSF) of 78N. When joint stressing loads remain below this stressing force level, the MCL will not damage in 95% of patients.

Table 1. Maximum stressing force of the Medial Collateral Ligament MCL with the standard deviation (SD)

Maximum stressing force MCL	*SD*	*Weakest stressing force MCL*	*Reference*
665N	75N	515N	Trent *et al*(1976)[16]
468N	33N	402N	Kennedy *et al*(1976)[14]
534N	85N	364N	Robinson *et al*(2004)[15]

Table 2. Measures and ratios of knee and lower leg. (C:B=D)

		Man	*Woman*
A	Mean Width of condyle (mm) [17]	51.6	45.6
B	Moment arm MCL (mm)	38.7	34.2
C	Length of lower leg (mm)[18]	429	389
D	Ratio of stressing force on the lower leg at ankle level and	11/1	11/1

Participants

The protocol was approved by Medical Ethical Commission. All patients signed informed consent. Twenty-one patients, eleven men and ten women (median 46 (range 18-73) years), that were scheduled for a routine knee arthroscopy in the day care centre in the Academic Medical Centre (AMC) in Amsterdam, Netherlands, were included. The median body mass index was 27 (range: 20.3-40.6) kg/m2. The mean medial leg length was 410 mm (SD: 20), the mean lateral leg length was 420 mm (SD: 20). Arthroscopy was performed for a meniscal lesion (n=11), ostheoarthritis (n=3), tendinitis of the right patella ligament (n=1), vascular pseudo-arthrosis (n=1), removal of free body (n=1), removal of plica mediopatellaris (n=1) or a diagnostic purpose (n=3). The level of knee (in)stability was tested with a Lachman test[19] (17), which demonstrated that all knees were stable.

The power analysis focused on achieving a sufficient dataset that reflected the variation in the entire population. During the pilot test, two researchers each stressed legs of 5 subjects laterally and medially, with the knee flexed at approximately 30°. Based on the standard deviation (21 N) from this pilot and an acceptable 95% confidence limit of +/- 9 N, the sample size was calculated to be 21 patients. This number would be sufficient to give a representative dataset of the existing variation in joint stressing during knee arthroscopy, and based on a comparison with the TMSF indicate if there is a need to set a threshold for safe joint stressing.

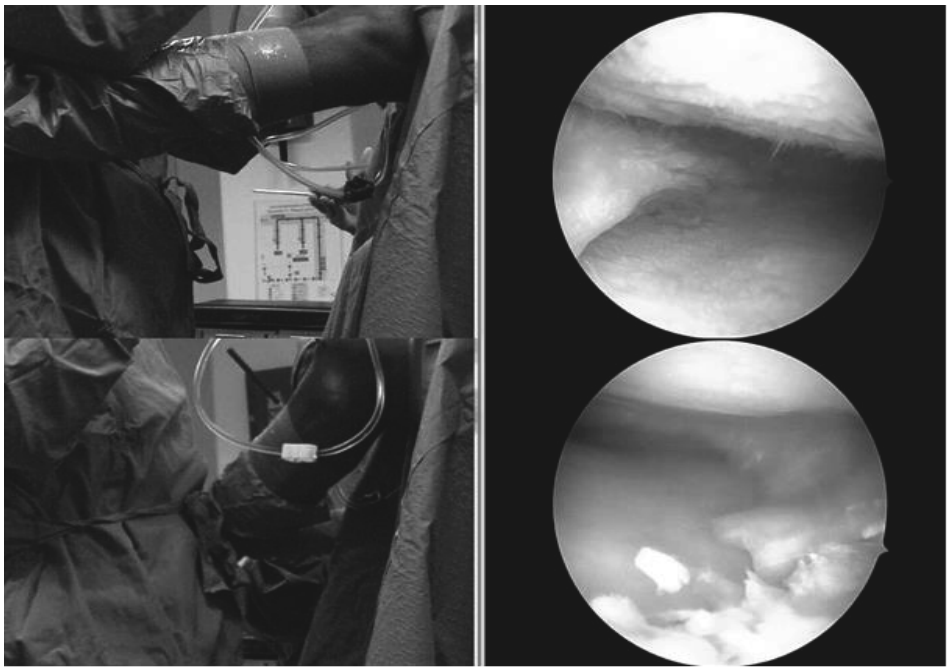


Figure 2 Pictures taken in the OR of the stressing motion of the lower leg in valgus (1A) and varus (2A) direction toincrease operating space inside the knee joint. Pictures 1B and 2B show the corresponding arthroscopic view of the joint space, respectively

Study design

The recording equipment consisted of a 1 load cell (LSB200 (L2357), Futek, Irvine CA, USA) which was integrated in a custom-made hip belt that could be placed underneath the sterile clothes of the surgeon (Figure 2). The load sensor was surrounded by a special construction to give an adequate pressure surface. The sensor was calibrated in that construction by loading the sensor with known loads to relate the voltage output to the force. The accuracy of the sensor was 0.1%.

Only forces in the varus and valgus direction are measured. The weight of the leg was compensated by the surgeon imposing the patient's heel in a U-shaped board (Figure 2). Furthermore, a dedicated data-acquisition-system was used consisting of a Notebook PC with custom-

made software developed VISIONDAQ (version 1.2, MTO, AMC), which was connected with a video server (GS-C4CQR Golden state Instrument co., Tustin, USA), the camera unit of the arthroscope and a separate camera (digital CCD camera, 21CW Sony CCD, Tokyo, Japan) (Figure 2). The data acquisition system enabled the synchronized acquisition of the force data channels and two digital video streams. The sample rate of the data channel is set at 500 Hz and that of the video stream at 25 Hz. Thus simultaneously, the arthroscopic view, the view on the flexion-extension of the knee joint and the stressing force were recorded.

Before surgery started, age, sex and the lower leg length were documented. Lower leg length was measured in two ways: a) the distance between the medial knee joint space and the medial malleolus and b) the distance between the lateral knee joint space and the lateral malleolus. During the operation, the surgeon placed the load sensor-belt around the hip and prepared for sterility thereby putting the surgical gown over the load sensor-belt. When the arthroscope was placed in the first portal (anterolateral portal) and the interior of the knee joint was visualized, the measurement of the data signals was started. Routine inspection of the knee joint started in the lateral compartment by stressing the knee joint in varus direction with the hip at the position of the load sensor. Subsequently, the load sensor-belt was rotated along the hip to the other side, and the medial compartment was inspected by stressing the knee joint in valgus direction again with the hip at the position of the load sensor.

Data processing

The force data was processed with MATLAB software (version 7.0.4.365 (R14), The Mathworks, Natick, USA). The first and last time frame the

measured stressing force were determined for both valgus stressing (opening of medial compartment) and varus stressing (opening of lateral compartment). A clear view was defined by a visible tibia plateau, femur condyle and meniscus. The first and last frames with a clear arthroscopic view on the knee joint space were selected as the first and last time frame, respectively. Within this set of measured forces, the 10 highest force values were automatically detected and averaged to compensate for noise (Figure 3). This average value was used as the maximum force value for each particular patient (Fmaxmed and Fmaxlat). The joint stress moment (Mmaxmed and Mmaxlat) was calculated by the force data multiplied with the lower leg length.

After recording the procedures, the arthroscopic view was used to determine the moment of adequate joint stressing and the external camera is used to determine the flexion-extension angle relative of the patient's lower leg relative to the upper leg. The force value at the start time frame with a clear arthroscopic view in the joint compartment was used to determine the minimal stressing force needed to open the joint compartment. For instance, in Figure 3, the start force value is the value at the first vertical dotted line, indicating the start of the measurement (Mstartmed and Mstartlat). The recordings with the separate camera allowed us to indicate the level of flexion of the knee during joint stressing.

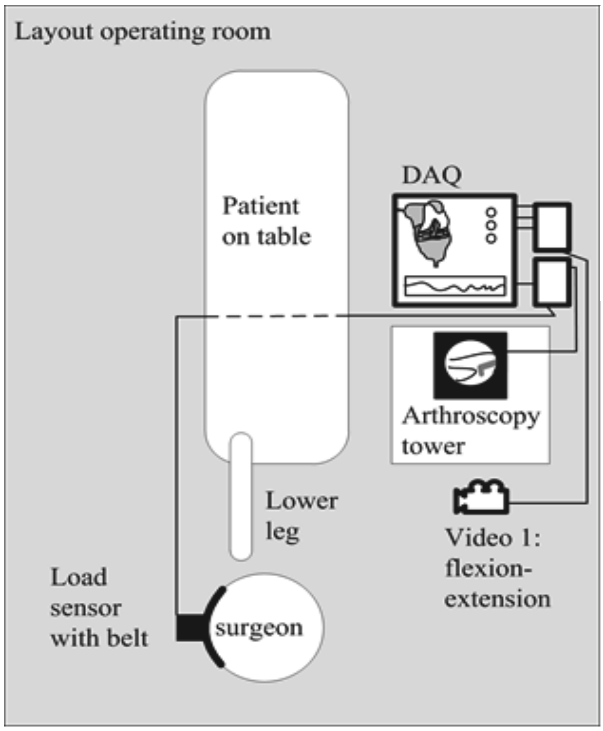


Figure 3: schematic drawing indicating the set up of the measurement equipment in the operation room: Video1, load sensor and data-acquisition system (DAQ)

Statistical analysis

The data were analyzed by use of SPSS version 12.0 (Chicago, IL) to determine if the set was normally distributed using the Kolmogorov-Smirnov tests were performed. In case of normality, data were described as means with standard deviations and comparison between varus and valgus stressing was performed using paired T-tests. A level of $p < 0.05$ was considered statistically significant.

Results

The force data showed a normal distribution. The mean medial lower leg length was 410 mm (SD=2 mm); the mean lateral lower leg length was 420 mm (SD=2 mm). Stressing was always performed with the knee slightly flexed. The average maximum stressing force was 55 N (SD=16

N) to open the medial compartment (valgus stressing) and 66 N (SD=36 N) to open the lateral compartment (varus stressing) (n.s.). As there was no significant difference the data points were pooled. This gave an average maximum stress force of 60 N with a standard deviation of 28 N, with a 95% interval of 8.5 N for the average joint stressing. Seven stressing forces exceeded 78 N (Figure 4). The confidence interval was 8.5 N. The mean first frame force was 43 N (SD=34 N) for varus stressing and 50 N (SD=33 N) for valgus stressing (n.s.). The first frame forces did not show significant differences either, therefore, data points were again pooled. This yielded a mean first force of 47 N (SD=34 N). The impact of leg length on applied stressing forces was taken into account by calculating the joint stressing moment. The average maximal stress moment was 25 Nm (SD=12 Nm), the average first frame moment was 20 Nm (SD=14 Nm).

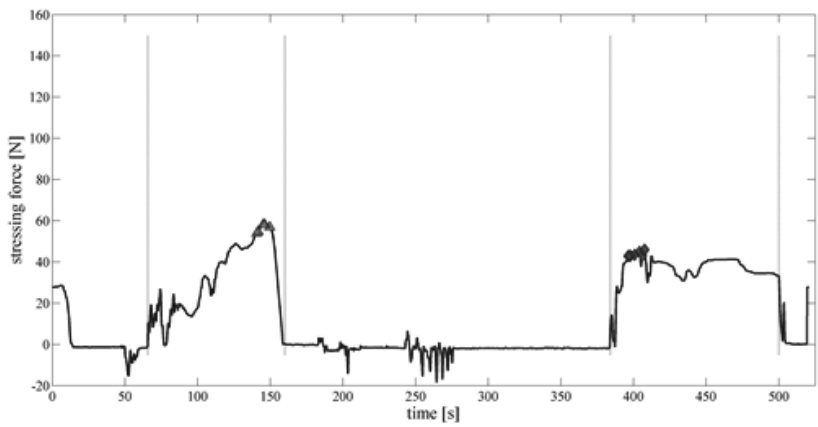


Figure 4 Raw force measurement data in time a particular patient. The triangles are the highest valgus data points and the diamonds are the highest varus data points. The start and end time of the measurement is shown as by four vertical dotted lines

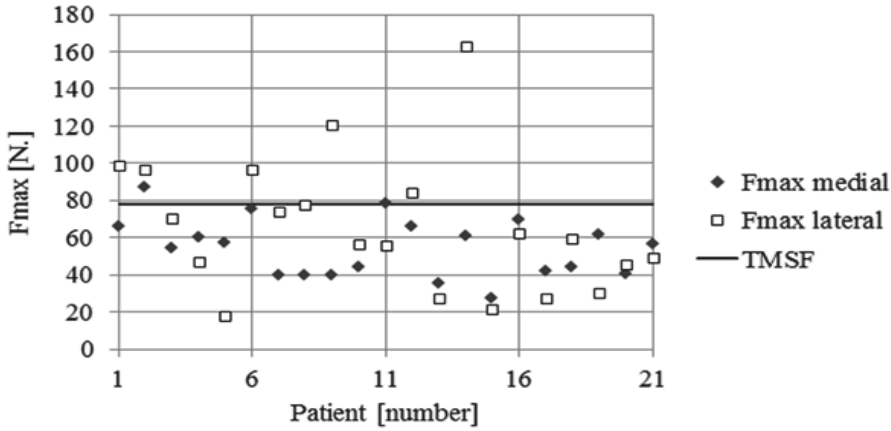


Figure 5 Scatter plot of medial and lateral maximal stressing forces per patient. The figure shows that 7 stressing forces exceed the Theoretical Maximal Stressing Force 78 N (straight line)

Discussion

The most important finding of the present study was that the overall mean maximum stressing force of 60 N (SD=28 N) was below the theoretically determined allowable force level (78 N). Additionally, the variation in stressing forces was relatively high. Maximum lateral stressing forces did not significantly differ from maximum medial stressing forces (n.s.). Also first frame lateral and medial stressing were not significantly different (n.s.). The variation of the stressing forces was high especially for lateral opening (standard deviation is 29% of the mean for medial opening and 54% for lateral opening). Variation was also high for the first frame stressing forces. The found variation is not surprising, since both multiple surgeons and patients were included in this study, and stressing forces are depended on both surgeon and patient. Surgeon's factors include factors such as strength, experience and prudence. Patient's factors include the bony attachment of the ligaments, which is dependent on the bone density on the site of bony attachment[16] ; the tensile strength and visco-elastic

properties of the ligaments, which is to a large extent dependent on age[14, 15]; and physical activity[20, 21]. The differences found between the leg lengths of patients were too small to be of any influence.

On average, the measured stressing forces did not exceed the theoretical maximum allowable stressing forces of 78 N. However, stressing forces in the OR were measured that did exceed the TMSF (Figure 5). Even in these cases though, no ligament failure was found, in contrast to earlier reports[14-16]. Several factors might contribute to this. Firstly, measurements of stressing forces of earlier studies were not performed in vivo, but on cadavers[14-16]. Those ligaments might be more prone to ligament failure than ligaments of our patients. Secondly, microscopic failure was not investigated; when determining maximum stressing forces, ligaments might have been stressed so much that microscopic failure occurred, but remained visibly intact[14].

There are several limitations to the study. First of all, the population (patients) differed from the population of most other studies on ligament properties (cadavers)[3, 10, 14]. Therefore, the calculated theoretical TMSF might not be directly applicable to real life settings. But since the maximum stressing force was on average less than 78 N, the TMSF was considered as a good initial value. Secondly, rotation was not included in the derivation of the TMSF. However, Hull et al. (1996) state that rotation of the lower leg with respect to the upper leg does not cause a significant difference in the size of forces on the collateral ligaments[22]. The effects of rotation of the lower leg were simulated with the 'knee joint simulator' program (MADYMO, TNC Road-Vehicles Research Institute, Delft, the Netherlands) and verified that indeed rotation of the lower leg does not cause significant differences to the medical collateral ligaments (MCL). Thirdly, patient's factors such as age, sex, lower leg length and BMI were

not discussed thoroughly. We realize that (one of) these factors might affect joint stressing forces. However, this was less relevant for our study purpose, which was to determine the variation in joint stressing forces to set threshold levels for safe tissue manipulation in a training environment. Moreover, the sample size is too small to give any significant results on the influence of these factors. Lastly, potential microscopic failure was not investigated. But as all knees were stable, patients did not have any post-operative complaints and all surgeons were very experienced, this type of microscopic failure did probably not occur.

Since the mean maximum stressing force was 60 N and the majority of the measured stressing forces was below 78 N (Figure 5), 78 N is considered a good initial value to train with. Conclusions can be drawn justly from the measured values, since our power analysis was correct, as the 95% confidence interval of the measured data was 8.5 N, and stayed below the 95% confidence interval limit of 9 N that was set acceptable. Thus, the data from the 21 participants are representative for the entire population. As no significant difference was found between varus and valgus stressing forces, different maximum values for varus and valgus stressing do not have to be set for training. However, since variation in stressing forces are high, it is also recommended to offer training cases where the complete range of stressing levels can be experienced. Some patients might have, for example, very strong ligaments, which require more stressing force. Simulators must have the opportunity to train these variations to enhance training. It is important to implement force feedback with safety levels in arthroscopic simulators when residents practice knee arthroscopies.

The results of this study should make training surgeons aware of the fact that teaching adequate joint stressing is not necessarily

straightforward, and that safe joint stressing levels are necessary for efficient training. Joint stressing training on an arthroscopic knee simulator could be the future manner of training. Repeated training will increase the competency level of residents and enhance patient safety as potential damage of the knee joint ligaments is minimized.

Conclusion

Since joint stressing can cause ligament failure, it is important to determine a maximum level of stressing force, and to train with this level. To remain in a safe deformation zone with no chance of tissue damage, the stressing force may not exceed the safe loading limits. Joint stressing levels are depended on both patient and surgeon, and relatively high variations in stressing forces were found. A maximum value of 78 N is a good initial value for residents and surgeons to stay below. However, since variation of joint stressing magnitudes is high, offering training cases on simulators where the complete range of stressing forces can be experienced is recommended.

References

[1] Meyer RD, Tamarapalli JR, Lemons JE: Arthroscopy training using a "black box" technique Arthroscopy 1993, 9:338-340.

[2] Nandi S, Parker R: Deep medial collateral ligament tear during knee arthroscopy. J Knee Surg 2012, 25:79-81.

[3] Tuijthof GJ, Horeman T, Schafroth MU, Blankevoort L, Kerkhoffs GM: Probing forces of menisci: what levels are safe for arthroscopic surgery. Knee Surg Sports Traumatol Arthrosc 2011, 19:248-254.

[4] Schmid RB, Wirz D, Gopfert B, Arnold MP, Friederich NF, Hirschmann MT: Intra-operative femoral condylar stress during arthroscopy: an in vivo biomechanical assessment. Knee Surg Sports Traumatol Arthrosc 2011, 19:747-752.

[5] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Br 2008, 90:494-9.

[6] Moody L, Waterworth AR, McCarthy AD, Harley P, Smallwood R: The feasibility of a mixed reality surgical training environment. Virtual Reality 2012, 12:77-86.

[7] Cannon WD, Eckhoff DG, Garrett Jr. WE, Hunter RE, Sweeney HJ: Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res 2006, 442:21-29.

[8] Heng PA, Cheng CY, Wong TT, Wu W, Xu Y, Xie Y, Chui YP, Chan KM, Leung KS: Virtual reality techniques. Application to anatomic visualization and orthopaedics training. Clin Orthop Relat Res 2006, 442:5-12.

[9] McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR: Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res 2006, 442:13-20.

[10] Safir O, Dubrowski A, Mirksy C, Backstein D, Carnahan H: What skills should simulation training in arthroscopy teach residents? Int J Cars 2008, 3:433-437.

[11] Tuijthof GJ, van Sterkenburg MN, Sierevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GM: First validation of the PASSPORT training environment for arthroscopic skills. Knee Surg Sports Traumatol Arthrosc 2010, 18:218-224.

[12] Mabrey JD, Cannon WD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Mevis H, Garrett WE, Poss R: Development of a virtual reality arthroscopic knee simulator. Stud Health TechnolInform 2000, 70:192-194.

[13] Zivanovic A, Dibble E, Davies B, Moody L, Waterworth A: Engineering requirements for a haptic simulator for knee arthroscopy training. Stud Health Technol Inform 2003, 94:413-418.

[14] Kennedy JC, Hawkins RJ, Willis RB, Danylchuck KD: Tension studies of human knee ligaments. Yield point, ultimate failure, and disruption of the cruciate and tibial collateral ligaments. J Bone Joint Surg Am 1976, 58:350-355.

[15] Robinson JR, Bull AM, Amis AA: Structural properties of the medial collateral ligament complex of the human knee. J Biomech 2005, 38:1067-1074.

[16] Trent PS, Walker PS, Wolf B: Ligament length patterns, strength, and rotational axes of the knee joint. Clin Orthop Relat Res 1976, 117:263-270.

[17] Mensch JS, Amstutz HC: Knee morphology as a guide to knee replacement. Clin Orthop Relat Res 1975, 112:231-241.

[18] http://dined.io.tudelft.nl/dined/nl.

[19] Smith BW, Green GA: Acute knee injuries: Part I. History and physical examination. Am Fam Physician 1995, 51:615-621.

[20] Laros GS, Tipton CM, Cooper RR: Influence of physical activity on ligament insertions in the knees of dogs. J Bone Joint Surg Am 1971, 90:494-499.

[21] Woo SL, Gomez MA, Sites TJ, Newton PO, Orlando CA, Akeson WH: The biomechanical and morphological changes in the medial collateral ligament of the rabbit after immobilization and remobilization. J Bone Joint Surg Am 1987, 69:1200-1211.

[22] Hull ML, Berns GS, Varma H, Patterson HA: Strain in the medial collateral ligament of the human knee under single and combined loads. J Biomech 1996, 29:199-206.

CHAPTER 8

The suitability of Global Rating Scales to monitor arthroscopic training progress

J.J. Stunt, G.M.M.J. Kerkhoffs, B. van Ooij, I.N. Sierevelt, M.U. Schafroth, C.N. van Dijk, J. Dragoo, G.J.M. Tuijthof

submitted

Abstract

Purpose: As developing arthroscopic skills is challenging and training time for residents is limited, arthroscopic skill competency of residents should be measured. Assessment tools, such as Global Rating Scales (GRS), have been developed for structured, objective feedback and to assess learning curves. The goal of this study is to assess known-groups and convergent validity of these scales, to evaluate the suitability of these scales to monitor training progress of residents.

Methods: Knee arthroscopies and ACL reconstructions performed by residents were supervised and assessed, using both GRS questionnaires. The estimates of the parameters were used to study the relationship between year of residency and each GRS score, and between the number of previously performed arthroscopies and each GRS score. Pearson correlation coefficient between GRS scores were calculated to measure convergent validity. A Bland-Altman plot with a paired t-test was constructed to evaluate the agreement between GRS I and II.

Results: Mixed model analysis revealed a significant increase ($p<0.001$) per year of residency on both GRSs (8.1 points (95% CI: 6.3-9.9) and 9.2 points (95% CI: 7.4-11.2) respectively). Significant increases per performed arthroscopy were also observed for both GRSs ($p<0.001$) (0.14 (95% CI: 0.09-0.18) and 0.13 points (95% CI: 0.08-0.2) respectively). Scores for ACL reconstructions were significantly lower ($p<0.001$) than for standard knee arthroscopies (12.5 and 13.0 points respectively, $p<0.001$). The Pearson correlation coefficient between GRS I and GRS II scores was high (0.94). The Limit of Agreement was 11 points.

Conclusion: GRS I and GRS II demonstrate sufficient construct validity. However, they seem not sufficiently sensitive and consistent to establish individual learning curves. Both scales are suitable to objectively evaluate global progress of residents in the operating room when acquiring arthroscopic skills, in particular on group level.

Key words: Global Rating Scales, objective assessment, learning curves, arthroscopic training

Introduction

Since arthroscopic surgery has several advantages compared with open surgery, it has become the most performed procedure in orthopaedic surgery[1-5]. However, as developing arthroscopic skills is challenging[6-9] and training time for residents is limited[8, 10-12], professional societies have requested arthroscopic skill competency of residents to be assessed to improve patient safety[8, 13]. Assessment of skills by expert surgeons is sensitive to the subjective opinion of the assessor, which might compromise fair judgment[14]. To overcome this issue, the formulation of criteria and proficiency levels for evaluation of arthroscopic skills is recommended.

Assessment tools for monitoring technical skills in the operating theatre, such as Global Rating Scales (GRS), have been developed for structured, objective feedback and to assess learning curves[15]. Previous research investigated whether GRS are valid, reliable tools to objectify resident performance in surgery; feasibility, face validity, content validity, construct validity and reliability have been demonstrated for various Global Rating Scales[13, 16-24]. Two Global Rating Scales that have been specifically proposed for feedback during arthroscopic training, are the Basic Arthroscopic Knee Skill Scoring System[6] and the Orthopaedic Competence Assessment Project[25]. Insel and co- workers developed a GRS to assess diagnostic knee arthroscopies and partial meniscectomies on cadaver knees (GRS I)[6]. This GRS has demonstrated validity, but only on cadavers, performing basic arthroscopic tasks[6]. Howells and co-workers combined the Orthopaedic Competence Assessment Project and the Objective Structured Assessment of Technical Skill (OSATS) (GRS II) to test arthroscopic simulator training on a bench-top knee simulator[25]. Validity of this GRS has however not yet been assessed.

Since both GRSs are not validated in a clinical context, the goal of this study is to assess validity of these scales during training of arthroscopic skills in the operating room on real-life patients. In the absence of a gold standard, both known-groups and convergent validity are investigated to assess the suitability of these scales to monitor training progress of residents.

Materials and methods

Participants

Twenty-eight orthopedic residents in four consecutive residency years (year 3 to 6) and ten experienced orthopedic surgeons were recruited at two institutions (the Stanford University School of Medicine and the Academic Medical Centre in Amsterdam) (Table 1).

Study design

After participants signed informed consent, all outpatient KA's and ACL reconstructions performed by each resident were supervised and assessed by one of the experienced surgeons, using both GRS questionnaires. Before each procedure, the resident's and supervisor's unique identifier code, type of operation, year of residency and number of previously performed arthroscopies were documented. Hundred-and-thirty-five KA's and 30 ACL reconstructions were included (Table 1). Residents performed on average six procedures, within a time frame of maximum a month.

Resident:	Date:	Completion time:
Supervisor:	Residency level:	
Operation:	Complexity: low/average/high	

Please circle the number (1-5) that describes the subject best.

1. Dissection
1- Appeared excessively hesitant, caused trauma to tissues, did not dissect into correct anatomical plane.
2-
3- Controlled and safe dissection into correct anatomical plane, caused minimal trauma to tissues.
4-
5- Superior and atraumatic dissection into the correct anatomical plane.

2. Instrument handling
1- Repeatedly makes tentative or awkward movements with instruments.
2-
3- Competent use of instruments, although occasionally appeared stuff or awkward.
4-
5- Fluid moves with instruments and no awkwardness.

3. Depth perception
1- Constantly overshoots target, slow to correct.
2-
3- Some overshooting or missing of target.
4-
5- Accurately directs instruments in the correct plane to target.

4. Bimanual dexterity
1- Noticeably awkward with non-dominant hand, poor coordination between hands.
2-
3- Uses both hands but does not maximize interaction between hands.
4-
5- Expertly uses both hands in complementary manner to provide optimum performance.

5. Flow of operation and forward planning
1- Frequently stopped operating or needed to discuss next move.
2-
3- Demonstrated ability for forward planning with steady progression of operative procedure.
4-
5- Obviously planned course of operation with effortless flow from one move to the next.

6. Knowledge of instruments
1- Frequently asked for the wrong instrument or used inappropriate instrument.
2-
3- Knew the names of most instruments and used appropriate instrument for the task.
4-
5- Obviously familiar with the instruments required and their names.

7. Efficiency
1- Many unnecessary, inefficient movements. Constantly changing focus or persisting without progress.
2-
3- Slow, but planned movements are reasonably organized with few unnecessary or repetitive movements.
4-
5- Confident, clear economy of movement and maximum efficiency.

8. Knowledge of specific procedure
1- Deficient knowledge, needed specific instruction at most operative steps.
2-
3- Knew all important aspects of the operation.
4-
5- Demonstrated familiarity with all aspects of the operation.

9. Autonomy
1- Unable to complete entire task, even with verbal guidance.
2-
3- Able to complete task safely with moderate guidance.
4-
5- Able to complete task independently

10. Quality of final product
1- Very poor.
2-
3- Competent
4-
5- Clearly superior.

Figure 1: GRS I Basic Arthroscopic Knee Skill Scoring System (BAKSSS)[6]

Resident:	Date:	Completion time:
Supervisor:	Residency level:	
Operation:	Complexity: low/average/high	

Please circle the number (1-5) that describes the subject best.

1. Follows protocol
1- Unsatisfactory.
2-
3- Adequate. Occasional need for guidance and help.
4-
5- Excellent adherence to agreed protocol. No prompts. No mistakes.

2. Handles tissue well
1- Careless. Potential to cause damage.
2-
3- Adequate. No tissue damage. Occasional need for increased care.
4-
5- Excellent tissue handling. Precise and delicate.

3. Appropriate and safe use of instruments
1- Dangerous. Risk to patient and assistant. Potential for damage to equipment.
2-
3- Adequate use of instruments and scope. Occasional guidance to ensure instruments remain within field of vision.
4-
5- Excellent use of instruments. Good control of arthroscope. Instruments constantly within field of vision.

4. Appropriate pace with economy of movement
1- Erratic pace and movements. Overly rushing or inappropriately slow.
2-
3- Adequate economy of movement. Majority of movements controlled and careful. Occasional erratic movement.
4-
5- Excellent fluidity and economy of movement. Procedure performed at appropriate pace without erratic movements.

5. Act calmly and effectively with untoward events
1- Unable to deal with adverse events. Panic and inability to respond.
2-
3- Remains calm. Remains safe. Takes advice from supervisor. Unable to cope independently.
4-
5- Excellent ability to cope with adverse events. Remains calm. Deals with complication independently.

6. Appropriate use of assistant
1- Fails to involve assistant appropriately. Resultant poor positioning. Poor rapport.
2-
3- Asks for appropriate joint position at appropriate times. Unable to suggest alternative positions to improve view/access.
4-
5- Excellent use of assistant. Good rapport. Able to constantly modify input of assistant to best advantage throughout procedure.

7. Communicates with scrub nurse
1- Inappropriate communication resulting in confusion or operative delay.
2-
3- Appropriate communication with scrub nurse. Occasional need for clarification from supervisor.
4-
5- Excellent rapport with scrub nurse. Clear and effective communication, maximising procedural efficiency.

8. Clearly identifies common abnormalities
1- Unable to identify common abnormalities. Confusion over basic anatomy.
2-
3- Adequate identification of common pathology. Occasional mistake. Unsure of precise classifications.
4-
5- Excellent knowledge of pathology of common abnormalities. Clear understanding of classification of injuries.

9. Protecting the articular surface
1- Inability to protect articular surface appropriately. Potential to cause damage.
2-
3- Awareness of need to protect articular surface. Adequate care taken. Occasional prompt from supervisor required.
4-
5- Excellent awareness of articular surfaces. High degree of care maintained throughout the procedure

Figure 2: Orthopaedic Competence Assessment Project (OCAP)[25]

Outcome measures

GRS I is a ten-item Global Rating Scale, derived from previously published and validated evaluation models to assess arthroscopic skills (Fig. 1). Items from these models were used to create a task-specific checklist and a global rating scale that together form the Basic Arthroscopic Knee Skill Scoring System (BAKSSS), a model specific to diagnostic knee arthroscopy and partial meniscectomy[6]. The checklist was not included in the current study. Items of the global rating scales include dissection, instrument handling, depth perception, bimanual dexterity, flow of operation, knowledge of instruments, to the knowledge of the specific procedure, autonomy, efficiency, and quality of the operative result[6].

For GRS II, nine of the fourteen Orthopaedic Competence Assessment Project (OCAP) criteria for diagnostic arthroscopy were selected (Fig. 2). Competences include following procedure protocol, handling of tissue, appropriate and safe use of instruments, appropriate pace with economy movement, calmness and effectiveness in dealing with untoward events, appropriate use of assistants, communication with scrub nurse, and identification of common abnormalities and protection of articular surface[25].

GRS I and II have similar domains, such as instrument handling, flow of operation, efficiency and autonomy. Both Global Rating Scales allow assessors to rate arthroscopic skills performance on each domain, using 5-point Likert scales with anchors at 1, 3, and 5 points. The anchor points have specific descriptions of the necessary requirements to receive the respective point values, which should help uniform assessment. Higher scores indicate better arthroscopic proficiency. Minimum and maximum GRS score for the GRS I are 10 and 50 points, respectively[6].

Minimum and maximum GRS score for GRS II are 9 and 45 points, respectively[25].

Firstly, known-groups validity will be investigated by determining the extent to which the GRSs can discriminate between levels of experience[26]. Secondly, convergent validity is investigated by determining whether the two GRSs correspond with one another, as they cover similar domains of arthroscopic skills[27]. To this end, knee arthroscopies (KA) and anterior cruciate ligament (ACL) reconstructions performed on real life patients will be used. If validity is shown, these GRS could be further developed into objective assessment tools to show individual training progress of residents.

Statistical analysis

Analysis was performed with SPSS 22© (SPSS Inc., Chicago, IL, USA). All 165 procedures were included in the analysis (Table 1). In order to compare GRS I and II, scores were normalized to a range from 0 to 100 points. Normality of the parameters was assessed using the Kolmogorov-Smirnov test, and the skewness and kurtosis of the sample (-2<z-value<2).

To account for correlated assessments within the residents (that is multiple assessments were performed per resident), a multilevel analysis was performed by use of mixed model analysis using a residual maximum likelihood (REML) approach (Table 2). The estimate of the parameters, as well as the standard error and confidence intervals were used to study the relationship between year of residency and each GRS score, and between the number of previously performed arthroscopies and each GRS score. Type of operation (KA of ACL reconstruction), year of residency and number of previously performed arthroscopies were entered as model factors, with GRS I and GRS II scores as dependent variables. Pearson

correlation coefficient between the scores of GRS I and GRS II were calculated to measure convergent validity. Correlation coefficients ≤ 0.35 were considered to represent weak correlations, 0.36 to 0.67 moderate correlations, and 0.68 to 1.0 high correlations, with coefficients ≥ 0.91 very high correlations[28, 29].

A Bland-Altman plot was constructed to evaluate the agreement between GRS I and II. The mean differences between GRS I and II scores against the absolute differences and limits of agreement (LoA) were calculated ($1.96*SD_{dif}$)[30]. A paired t-test was performed to assess a systematic difference between the two scales. P-values ≤0.05 were considered statistically significant.

Table 1. An overview of the number of residents per residency year, and the total number of performed procedures by the residents within one residency year

Year of Residency	Nr of residents	Total nr of procedures
3	8	32
4	6	36
5	8	45
6	6	52

Results

Non-normalized GRS sum scores varied between 19 and 50 points for GRS I and between 18 and 45 points for GRS II. In Table 2, results of the mixed model analysis are described. The parameters can be interpreted as the constant (intercept) and the coefficients or slopes (estimates) of the independent variables. Mixed model analysis revealed a statistically significant increase ($p<0.001$) per year of residency on both GRS I and II, with values of 8.1 points (95% CI: 6.3-9.9) and 9.2 points (95% CI: 7.4-11.2) respectively. Significant increases per performed arthroscopy were

also observed for both GRSs ($p<0.001$), with values of 0.14 (95% CI: 0.09-0.18) and 0.13 points (95% CI: 0.08-0.2) for GRS I and II respectively. Furthermore, scores for ACL reconstructions were significantly lower ($p<0.001$) than for standard knee arthroscopies (12.5 and 13.0 points for GRS I and II respectively, $p<0.001$) (Table 2).

Normalized GRS sum scores varied between 40 and 100 points for GRS I and between 38 and 100 points for GRS II. The scores did not differ significantly ($p=0.19$), with mean normalized scores of 70.8 (SD is 14.9) for GRS I and 71.3 (SD is 16.2). The Pearson correlation coefficient between the normalized GRS I and GRS II scores was high (0.94). The calculated LoA was 11 points, resulting in a lower limit of -11.6 and an upper limit of 10.4 (Fig. 3).

Table 2. Mixed model analysis showing the effects of year of residency, number of previously performed procedures and type of operation on scores for GRS I and II. All estimates are significant ($p<0.001$). The intercept can be interpreted as the mean of the outcome when all independent variables are zero. The estimates can be interpreted the same way as the estimates (coefficients) of predictors in a linear regression. GRS scores increase with 8.1 and 9.3 respectively per year of residency, and with 0.14 and 0.13 points respectively per number of previously performed arthroscopies. Scores for ACL-procedures are on average 12.5 and 13.0 points lower that KA scores.
s.e.: standard error; **CI**: confidence interval

	GRS 1	GRS II
	Estimate (95% CI)	**Estimate (95% CI)**
Intercept	34.4 (25.7-43.1)	29.4 (20.1-38.7)
Year of residency	8.1 (6.3-9.4)	9.3 (7.4-11.2)
No of arthroscopies	0.14 (0.1-0.2)	0.13 (0.1-0.2)
ACL-procedure	-12.5 (-17.2- -7.8)	-13.0 (-17.8- -7.8)

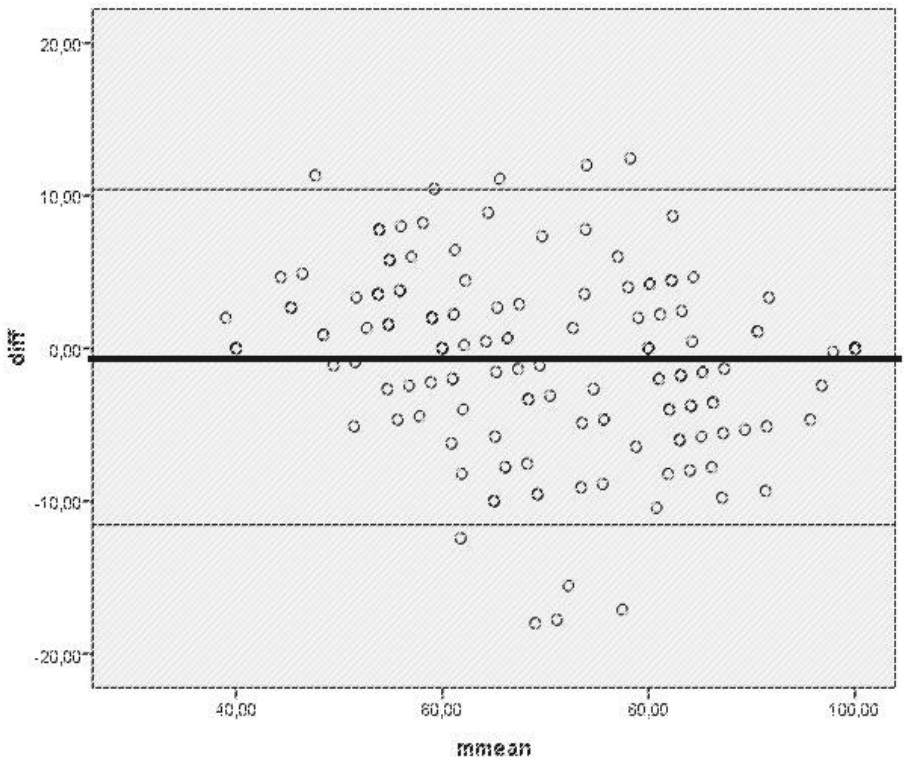


Figure 3: Bland–Altman plot comparing the scores of GRS I and GRS II. The solid line represents the mean difference between the two (-0.57), and the dotted lines represent the upper (10.4) and lower (-11.5) limits of agreement (mean ± 1.96 SD)

Discussion

This study investigated if the proposed GRSs show construct validity, more specifically, known-groups validity and convergent validity. With the available sample size, the study demonstrated that both GRS I and II were able to discriminate based on year of residency or number of arthroscopies, supporting known-groups validity. Convergent validity of the studied Global Rating Scales was supported by a high Pearson correlation coefficient. The Bland and Altman plot demonstrated that the average discrepancy between GRS I and II was small (close to zero),

indicating that there was no systematic difference between the two scales. However, the limits of agreement had higher values then the estimated differences per residency year or arthroscopic intervention: the standard error was 11, meaning that only differences larger than 11 points between the GRSs can be interpreted as actual difference when used on individual level. This emphasizes the question whether the GRSs are reliable outcome measures and if they are suitable for performance monitoring on individual level.

Besides year of residency, GRS scores are influenced by other factors, which can account for variability. One of these factors is the complexity of the type of procedure: ACL reconstructions are more complex than knee arthroscopies, which was reflected in the significantly lower GRS scores for scores for ACL reconstructions. Other factors influencing GRS scores are the complexity of the joint (depending on the anatomy or the severity of the condition of the patient), inter-observer differences between supervisors and the moment of the day or of the week at which the procedure was performed. Thus, as GRS score is determined by other factors additional to level of experience, in particular type and complexity of procedure, a standardized setting is required when using the GRS to measure competence.

The GRSs did not show floor or ceiling effects. None of the participants scored lower than 19 (GRS I) and 18 (GRS II) points, whereas the minimum values of the scales are 10 and 9, respectively. Moreover, no item was scored below two points. This can be attributed to the range of residency years that was included: residents were selected from their third year of residency, as they start than with their specialization in orthopaedic surgery in the Netherlands. Hence, none of the residents participating in the current study was a completely untrained

novice, as opposed to the original studies that also included participants with none or very little training[6, 25]. Our results indicate that the GRSs can be used for the entire duration of the residency curriculum.

The current study has limitations. Firstly, residents were assessed by one supervisor, implicating that inter-observer reliability could not be assessed. Secondly, supervisors were not blinded and thus were aware of the level of training of the residents. Thirdly, supervisors were not specifically trained to use the two Global Rating Scales of study. Ideally, all supervisors should have been trained with distinct examples, and multiple observers, who were blinded from the identity of the residents, should have performed the scoring. This would have increased consensus and objectivity. Unfortunately, this was logistically difficult to arrange. Vogt showed that knowing the identity of the resident does not significantly affect scoring[31]. Moreover, the study design was similar to other studies showing the potential of Global Rating Scales to objectively evaluate arthroscopic skills[6, 25, 32]. Therefore, we expect that these limitations will have marginally influenced the results.

As competency-based education is becoming more important in arthroscopic training [33], objective tools for assessment and performance monitoring of orthopedic residents need to be validated. The current study showed known-groups validity and convergent validity for Global Rating Scales. However, the results also suggest the scales do not seem to be sufficiently sensitive and consistent to monitor individual learning curves and progress of a trainee over a short period of time. Rather, they are suitable to objectify and assess general arthroscopic performance on group level in a structured way. Moreover, they can be applied in a research setting; to study differences on group level and to perform sample size calculation required to detect significant differences between different

levels of experience. Lastly, as feedback on performance is known to improve the learning process[34-37], and the structure of the GRSs allows feedback per skill domain, GRSs can also be valuable as educational tools.

Conclusion

The Basic Arthroscopic Knee Skill Scoring System (GRS I) and the Orthopaedic Competence Assessment Project (GRS II) demonstrate sufficient construct validity when performing knee arthroscopy or ACL reconstruction. However, they seem not sufficiently sensitive and consistent to establish individual learning curves. Both scales are suitable to objectively evaluate global progress of residents in the operating room when acquiring arthroscopic skills, in particular on group level.

Acknowledgement:

The authors are grateful to Joris Lansdaal and Mikel Reilingh for their contribution to enable the GRS assessments, and Robert van de Broek for his contribution to the set up and the execution of this study. We also want to thank the Slotervaart Hospital in Amsterdam, the Tergooi Hospital in Hilversum for their cooperation.

References

[1] Hobby J, Griffin D, Dunbar M, Boileau P: Is arthroscopic surgery for stabilisation of chronic shoulder instability as effective as open surgery? A systematic review and meta-analysis of 62 studies including 3044 arthroscopic operations. J Bone Joint Surg 2007, 89:1188-1196.

[2] Kim S, Bosque J, Meehan JP, Jamali A, Marder R: Increase in outpatient knee arthroscopy in the United States: a comparison of National Surveys of Ambulatory Surgery, 1996 and 2006. J Bone Joint Surg 2011, 93:994-1000.

[3] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner CD, Garrett WE, Olson T, Warme WJ, Nicandri GT: The Arthroscopic Surgical Skill Evaluation Tool (ASSET). Am J Sports Med 2013, 41:1229-1237.

[4] Poss R, Mabrey JD, Gillogly SD, Kasser JR, Sweeney HJ, Zarins B, Garrett WEJ, Cannon WD: Development of a virtual reality arthroscopic knee simulator. J Bone Joint Surg Am 2000, 82-A(10):1495-1499.

[5] Zhang YZ: Innovations in Orthopedics and Traumatology in China. Chin Med J (Engl) 2015, 128:2841-2842.

[6] Insel A CB, Leger R, Arciero R, Mazzocca AD: The development of an objective model to assess arthroscopic performance. J Bone Joint Surg Am 2009, 91:2287-95.

[7] O'Neill PJ, Cosgarea A, Freedman JA, Queale WS, McFarland EG: Arthroscopic proficiency: a survey of orthopaedic sports medicine fellowship directors and orthopaedic surgery department chairs. Arthroscopy 2002, 18:795-800.

[8] Hodgins JL, Veillette C: Arthroscopic proficiency: methods in evaluating competency. BMC Med Educ 2013, 13:613-621.

[9] Tuijthof GJ, Visser P, Sierevelt IN, Van Dijk CN, Kerkhoffs GM.: Does perception of usefulness of arthroscopic simulators differ with levels of experience? Clin Orthop Relat Res 2011, 469:1701-1708.

[10] Kairys JC, McGuire K, Crawford AG, Yeo CJ Cumulative operative experience is decreasing during general surgery residency: a worrisome trend for surgical trainees? J Am Coll Surg 2008, 206:804-811.

[11] Peabody T, Nestler S, Marx C, Pellegrini V Resident duty-hour restrictions-who are we protecting? AOA critical issues. J Bone Joint Surg Am 2012, 94:e131: 1-7.

[12] Sonnadara RR, van Vliet A, Safir O, Alman B, Ferguson P, Kraemer W, Reznick Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. Surgery 2011, 149:745-749.

[13] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A: The Arthroscopic Surgical Skill Evaluation Tool (ASSET). Am J Sports Med 2013, 41:1229-1237.

[14] Mabrey JD, Gillogly S, Kasser JR, Sweeney HJ, Zarins B, Mevis H, Garrett WE Jr, Poss R, Cannon WD: Virtual reality simulation of arthroscopy of the knee. Arthroscopy 2002, 18: E8.

[15] Ahmed K, Darzi A, Athanasiou T, Hanna GB.: Observational tools for assessment of procedural skills: a systematic review. Am J Surg 2011, 202:469-480.

[16] Bann S, Davis IM, Moorthy K, Munz Y, Hernandez J, Khan M, et al: The reliability of multiple objective measures of surgery and the role of human performance. Am J Surg 2005, 189:7477-52.

[17] Bann S, Kwok KF, Lo CY, Darzi A, Wong: Objective assessment of technical skills of surgical trainees in Hong Kong. Br J Surg 2003, 90:1294-1299.

[18] Ezra DG, Aggarwal R, Michaelides M, Okhravi N, Verma S, Benjamin L: Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents. Ophthalmology 2009, 116:257-262.

[19] Faulkner H, Regehr G, Martin J, Reznick R: Validation of an objective structured assessment of technical skill for surgical residents. Am J Surg 1996, 71:1363-1365.

[20] MacRae H, Regehr G, Leadbetter W, Reznick RK: A comprehensive examination for senior surgical residents. Am J Surg 2000, 179:190-193.

[21] Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA: Ceiling effect in technical skills of surgical residents. Am J Surg 2004, 188:294-300.

[22] Kim J, Neilipovitz D, Cardinal P, Chiu M: A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). Simul Healthc 2009, 4:6-16.

[23] Ilgen JS, Ma IW, Hatala R, Cook DA: A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ 2015, 49:161-173.

[24] Middleton RM, Baldwin MJ, Akhtar K, Alvand 4, Rees JL: Which Global Rating Scale? A Comparison of the ASSET, BAKSSS, and IGARS for the Assessment of Simulated Arthroscopic Skills. J Bone Joint Surg Am 206, 98:75-81.

[25] Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL: Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Br 2008, 90:494-499.

[26] Hattie J, Cooksey RW: Procedures for Assessing the Validities of Tests Using the "Known-Groups" Method. Appl Psy Meas 1984, 8:295-305.

[27] Campell DT, Fiske DW: Convergent and discriminant validation by the multitrait-multimethod matrix. Psychological Bulletin 1959, 56:81-105.

[28] Mason RO, Lind DA, Marchal WG: Statistics: An introduction. New York: Harcourt Brace Jovanovich, 1983.

[29] Weber JC, Lamb DR: Statistics and Research in Physical Education. St. Louis: CV Mosby Co, 2012.

[30] Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986, 1:307-310.

[31] Vogt VY, Givens VM, Keathley CA, Lipscomb GH, Summitt Jr RL: Is a resident's score on a videotaped objective structured assessment of technical skills affected by revealing the resident's identity? Am J Obstet Gynecol 2003, 189:688-691.

[32] van Heest A, Kuzel B, Agel J, Putnam M, Kalliainen L, Fletcher J: Objective structured assessment of technical skill in upper extremity surgery. J Hand Surg Am 2012, 37:332-337.

[33] Atesok K, Mabrey JD, Jazrawi LM, Egol KA: Surgical simulation in orthopaedic skills training. J Am Acad Orthop Surg 2012, 20:410-422.

[34] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35:883-888.

[35] Price AJ, Erturan G, Akhtar K, Judge A, Alvand A, Rees JL: Evidence-based surgical training in orthopaedics Bone Joint J 2015, 97-B:1309-1315.

[36] Harewood GC, Murray F, Winder S, Patchett S: Evaluation of formal feedback on endoscopic competence among trainees: the EFFECT trial. Ir J Med Sci 2008, 177:253-256.

[37] O'Connor A, Schwaitzberg SD, Cao CG: How much feedback is necessary for learning to suture? Surg Endosc 2008, 22:1614-1619.

CHAPTER 9

DISCUSSION

High-fidelity arthroscopic simulators are becoming important training tools for orthopedic surgery[1]. As developments continue, the need increases to integrate simulation into orthopedic training programs and to validate simulation curricula. Simulator training has been found to improve arthroscopic skills and shorten the learning curve for acquirement of basic arthroscopic skills[2-4]. However, there are still major barriers to the use of simulators in orthopedic education, as attempts to yield validity data have been partially successful[5].

The establishment of simulators as a standard training tool for arthroscopic skills relies on proper validation of simulators and performance monitoring tools. Consequently, the aim of this thesis was three-fold:

- To study the validity of arthroscopic simulators, the effect of simulator training and the transferability of arthroscopic skills;
- To study and validate metrics and complementary thresholds for safe and non-damaging tissue handling;
- To investigate the validity and utility of global rating scales for objective assessment and monitoring of arthroscopic performance.

Using a standardized protocol, we demonstrated basic levels of validity for the simulators of study, and revealed the shortcomings of the simulators, and applied methods and procedures. We gathered the information necessary for the development of a large-scale RCT on transferability of arthroscopic skills. Furthermore, the importance and impediments of incorporating performance metrics and thresholds in simulation environments were stressed, and we argued that global rating scales are appropriate for the assessment of general arthroscopic performance, but not for the establishment of individual learning curves.

When studying arthroscopic skills acquisition, it is preferable to include every joint. Narrowing down to the knee joint makes sense, as knee arthroscopies are one of the most performed arthroscopic procedures, and the number of annually performed knee arthroscopies will continue to grow, due to the aging population[6, 7]. It sets a representative example for arthroscopic surgery, and the results of this study are useful for every area of arthroscopic surgery.

Validity of arthroscopic simulators

Effect of simulator training and transferability of arthroscopic skills

As technology advanced, simulators have become more sophisticated, realistic and affordable. However, technological improvement does not validate the use of simulators in medical training. Our review (Chapter 3) revealed that high-quality research on the different aspects of validity of current commercially available medical simulators is marginal; this counts especially in the field of arthroscopic surgery. Lack of validation bears a risk of stimulating incorrect skills training, which would compromise patient safety. Moreover, heterogeneity among studies makes it difficult to compare the value and applicability of various simulator environments available to arthroscopic educators.

Recently, evidence for the educational value of arthroscopic simulators has been accumulating and numerous studies have provided verification of simulator validity. However, recent reviews on arthroscopic simulation agree that studies are methodologically weak, heterogeneous, and that evidence for higher levels of validity still misses[8-12]. We have overcome several limitations in our validation studies: in line

with a protocol development in our group[13], we have continued to use that protocol to demonstrate face and construct validity for an updated version of a hybrid model, and a new virtual reality simulator (Chapters 4 and 5). The strength of these studies was the use of a standardized protocol and validated outcome measures, effectuating more reliable and better interpretable results. Face validity, and construct validity between novices and experts were demonstrated.

Like other studies on construct validity[10, 12], we were not able to distinguish between novices and intermediates and intermediates and experts. However, we should reconsider to what extent demonstrated differentiation from the intermediate group is a prerequisite for a simulator to be of value: the most important asset of arthroscopic simulators is their deployment as a means to prepare residents for the operating room, shifting the first part of the learning curve from the operating to the simulation environment. This ensures a sufficient level of competency before residents continue their training on actual patients. Simulator training may also be continued in the course of the orthopedic curriculum, contributing to updating and maintaining skills once proficiency has been achieved. But the most prominent place for simulator training should be at the start of orthopedic training, as this is most beneficial to training efficiency and patient safety. From this point of view, the requisite for a simulator to differentiate between adjacent groups in terms of experience is less critical.

The question of transferability of arthroscopic simulators is still practically unanswered[9, 10, 12, 14]. The most appropriate study design to investigate this is a randomized controlled trial, as this set up reduces spurious causality and bias. Our pilot study (Chapter 6) demonstrated positive trends with respect to transfer validity. However, the most

important result of this study is the information that has become available about appropriate methods, procedures and tools to set up a large scale RCT: a sample size of 25 subjects per study arm and a training period with at least five training sessions, is essential to establish a significant effect. These conditions are challenging, organizationally speaking, as they imply sufficient availability of simulators, and objective (blinded) and trained assessors, and appeal to the flexibility of participants and their working schedules. Moreover, there appear to be limitations of existing assessment tools and performance metrics to evaluate competency and progression of novices: firstly, global rating scales are not sufficiently sensitive to monitor the individual progression of novices in a relatively short training period. Secondly, existing validated outcome measures, such as task time, are too difficult for novices. This calls for more sensitive assessment methods, such as time-action analysis, and modified outcomes measures or new parameters to be developed, appropriate for the assessment of novices. Our findings provide guidance for the development of the first high-quality large scale RCT that can offer clarity about the effectiveness and transfer validity of simulator training.

Future perspectives for validity studies

Improved studies for available simulators, every update of an existing simulator, and every new simulator are required. Forthcoming studies should focus on a proper study design that overcomes limitations of existing studies and enables to investigate the effect of simulator training in an unbiased way. This includes validated outcome measures, sufficient power, appropriate study groups, extended study periods and adequate assessment methods. Such improvements will only be feasible with financial support of funding organizations for scientific research and if

researchers and manufacturers team up and collaborate to develop protocols and study designs to test simulators simultaneously with large groups of participants. This way, validated and improved simulation environments can be offered, which would take away a major concern of orthopedic teaching staff.

Further studies on transfer validity should be based on randomized controlled trials to demonstrate that simulator training translates into better arthroscopic skills on real patients. Our pilot study is a first move to a solid, large-scale RCT. Future research must also determine the number and timing of training sessions needed to reach an appropriate level of performance and to transfer skills from simulation to the OR, and whether there is a ceiling effect associated with the benefit of simulator training for the acquisition of arthroscopic skills. Lastly, international standardization of validation protocols are required, to compare simulators, acquisition, development and retention of skills, facilitate multi-center trials and to set an international standard for the requirements of simulators and skill competency. This means that hospitals and universities will have to cooperate on international level to set up and carry out multi-centre studies on simulator validation and to implement the results.

As the appropriateness and value of a specific simulator also depends on the training goals, educators should adjust the choice for a simulator to the type of skills that need to be trained, and the complexity of these skills. The difficulty level of skills affects the minimum required level of validity of a simulator as well: a high-fidelity simulator with a high validity level is of more importance for training of (minimal invasive) surgical skills than for training of patient examination skills, as surgical skills are more complex and may cause severe consequences for

the patient when trained incorrectly. In addition, it should be noted that simulator training is no substitute for cadaver training or training in the operating room. Thus, when a minimum required level of competency has been reached, training should be continued on cadavers and in the operating room.

Validity of performance monitoring tools

Thresholds for safety performance metrics

The focus so far has been mainly on metrics for performance efficiency. We were the first to complement performance efficiency with performance safety; a necessary follow-up, as both type of metrics are crucial in order to reflect performance optimally and to provide adequate feedback for the improvement of competency in arthroscopic skills. We investigated maximum tolerable forces for safe tissue manipulation (Chapter 7). Tissue properties of ligaments were determined from earlier experiments[15-17]. Our findings have demonstrated that teaching adequate force magnitudes is not necessarily straightforward, because stresses or loads cannot directly be seen. We have determined thresholds for save metrics and shown how to implement force feedback with safety levels in arthroscopic simulators so that residents can use the tactile feedback to experience the complete range of stressing levels and to train safe tissue manipulation when practicing knee arthroscopies. Integration of safety metrics and thresholds in simulation has been limited until now[18, 19], as it is difficult to reach a sufficient level of realism. Moreover, development and implementation of metrics and thresholds is complicated and time-consuming: every metric must be defined, for each metric or task optimal performance needs to be determined, and thresholds for absolute

maximum forces need to be determined and compared with theoretical maximum forces.

Future perspectives for the development of safety performance metrics and thresholds

The results of Chapter 7 demonstrated that effective learning could benefit from monitoring the force and the amount of manipulation of the knee exerted during training. Unfortunately, there is a lack of built-in safety performance metrics and complementary thresholds in currently available simulators. As learning to safely manipulate tissue is mandatory in the beginning of a training process, research on performance metrics should focus more on safety- and force-related parameters. More quantitative data need to be collected to design and validate metrics and to support their applicability with evidence-based thresholds to set limits for safe tissue manipulation and performance efficiency. This is definitely required for the future, if physical or virtual trainers are to be deployed as standard devices to train knee arthroscopies and to enable practicing of safe tissue manipulation[17, 20].

Evaluating and monitoring arthroscopic performance

In response to the need for more holistic assessments tool for performance monitoring, global rating scales for arthroscopic skills have been developed. In Chapter 8, we found that the OCAP and BAKKKS are useful assessment tools, as they demonstrate construct validity, are easy to implement in residency curricula, not costly and usable for either a simulation environment or the operating room[21]. Although it was argued repeatedly that GRSs are feasible tools for the assessment of an individual

learning process[21-23], our results in Chapter 6 and 8 evidently contradict this as it was shown that neither the OCAP, nor the BAKKKS or the ASSET was acceptably sensitive and consistent to follow individual learning and progress of a trainee over a short period of time. Consistency can be increased by training assessors to use global rating scales, by performing assessment by multiple assessors at the same time, and by blinding assessors from the identity of residents. This can be achieved by evaluating performances at a remote time and location with video recordings. For accurate assessment of specific tasks, sensitive assessment tools, such as performance metrics that can be incorporated in the simulator, are more appropriate than global rating scales. Global rating scales are best suited for the assessment of general arthroscopic performance on group level, to study differences on group level and to perform sample size calculation required to detect significant differences between different levels of experience.

Future perspectives for research on Global Rating Scales

Today, GRSs for arthroscopic skills have been validated for construct, content and concurrent validity as well as internal consistency, inter-rater and test-retest reliability[20, 24-32]. However, none of the study designs for validation are the same, thus one-to-one comparison is not possible. Moreover, the reported outcomes are all surrogate outcomes focusing on technical skills, instead of patient-related outcomes. Thus, standardized research is required to investigate how assessment can be embedded in orthopedic training curricula to support objective skills monitoring and effective learning of arthroscopic skills.

Other factors influencing skills acquisition

Evaluating the effect of simulator training is complex, as more factors than training alone can affect arthroscopic skills acquisition and performance. Variations in innate skills for psychomotor tasks were suggested to cause variations in arthroscopic skills performance[33]. Researchers have stated that computer video game experience correlates with VR simulator performance[34-36], and some even fear that VR simulators would evaluate a subject's talent for video games rather than their surgical skills[37]. However, other studies have contradicted this hypothesis[38-40]. In Chapter 6 we saw some effect of video game experience, but this effect was not significant. Sex differences have been found to exist with respect to simulator performance, with men showing higher interest[41] and performance then women[40, 42, 43]. We did not find significantly higher scores for men then for women, neither did Gomoll and co-workers[37]. Found sex differences were possibly due to a spurious relationship between sex and arthroscopic performance: in our validity studies we found men to be more experienced with video games then women, simulating an association between sex and arthroscopic performance. Furthermore, feedback from the instructor has been repeatedly found to have a positive effect on arthroscopic performance [33, 44, 45]. This also applies to preoperative warming-up[46-48], although this was again disproved[49]. Obviously, environmental distractions, negative stress coping and lack of self-efficacy, alcohol intake and sleep deprivation before a surgical task negatively affect performance[19, 50, 51].

In summary, there are numerous factors that are thought to affect arthroscopic performance. As there is yet no complete understanding of the extent of the influence of these factors, this should be studied more extensively, to ensure that these factors are adequately accounted for in

validity studies and during simulator training. Moreover, influencing factors should be introduced in simulation (as far as possible) to represent the operating theater more accurately[11].

In conclusion

As time-efficient, effective and safe methods of training are essential, the role of simulation in orthopedic surgical training will very likely continue to expand and objective and valid tools for assessment and performance tracking will also become more important. Sophisticated high-fidelity arthroscopic simulators have great potential as effective training modalities, offering standardized, controlled and adaptable environments to teach orthopedic residents arthroscopic skills and procedures, without jeopardizing patient safety and with an opportunity for feedback and assessment[21, 52]. Both sensitive and generic assessment tools should have a prominent role in evaluating the effect of simulator training and improvement of arthroscopic performance. Sensitive tools, such as performance metrics are specifically focused on patient safety and suitable to reflect individual learning curves. More generic tools, such as global rating scales enable overall performance assessment or provide a global impression of performance. A future approach to arthroscopic training should be a combination of simulator- and traditional training, canceling out the drawbacks of each method, while improving arthroscopic performance and competence, ultimately leading to increased patient safety.

References

[1] Stefanidis D, Scerbo MW, Montero PN, Acker CE, Smith WD: Simulator Training to Automaticity Leads to Improved Skill Transfer Compared With Traditional Proficiency-Based Training A Randomized Controlled Trial. Annals of Surgery 2012, 255:30-37.

[2] Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, Ramel S, Smith CD, Arvidsson D: Proficiency-based virtual reality training significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg 2007, 193:797–804.

[3] Kairys JC, McGuire K, Crawford AG, Yeo CJ: Cumulative operative experience is decreasing during general surgery residency: a worrisome trend for surgical trainees? J Am Coll Surg 2008, 206:804-811.

[4] Rosen J, Solazzo M, Hannaford B, Sinanan M: Objective laparoscopic skills assessments of surgical residents using Hidden Markov Models based on haptic information and tool/ tissue interactions. Stud Health Technol Inform 2001, 81:417-423.

[5] Pedowitz RA, Marsh JL: Motor skills training in orthopaedic surgery: a paradigm shift toward a simulation-based educational curriculum. J Am Acad Orthop Surg 2012, 20(7):407-409.

[6] Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ: Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 2005, 19:1523-1532.

[7] Issenberg SB, McGaghie WC, Petrusa ER, Lee GD, Scalese RJ: Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. Med Teach 2005, 27:10-28.

[8] Aim F, Lonjon G, Hannouche D, Nizard R: Effectiveness of Virtual Reality Training in Orthopaedic Surgery. Arthroscopy 2015, 32:224-232.

[9] Frank RM, Erickson B, Frank JM, Bush-Joseph CA, Bach Jr BR, Cole BJ, Romeo AA, Provencher MT, Verma NN: Utility of modern arthroscopic simulator training models. Arthroscopy 2014, 30:121-133.

[10] Slade Shantz JA, Leiter JR, Gottschalk T, MacDonald PB: The internal validity of arthroscopic simulators and their effectiveness in arthroscopic education. Knee Surg Sports Traumatol Arthrosc 2014, 22(1);33-40

[12] Vaughan N, Dubey VN, Wainwright TW, Middleton RG: A review of virtual reality based training simulators for orthopaedic surgery. Med Eng Phys 2015, 38:59-71.

[13] van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelman J: Objective assessment of technical surgical skills. Br J Surg 2010, 97:972-987.

[14] Modi CS, Morris G, Mukherjee R: Computer-simulation training for knee and shoulder arthroscopic surgery. Arthroscopy 2010, 26:832-40.

[15] Kennedy JC, Hawkins RJ, Willis RB, Danylchuck KD: Tension studies of human knee ligaments. Yield point, ultimate failure, and disruption of the cruciate and tibial collateral ligaments. J Bone Joint Surg Am 1976, 58:350-5.

[16] Robinson JR, Bull AM, Amis AA: Structural properties of the medial collateral ligament complex of the human knee. J Biomech 2005, 38:1067-1074.

[17] Trent PS, Walker PS, Wolf B: Ligament length patterns, strength, and rotational axes of the knee joint. Clin Orthop Relat Res 1976, 117:263-270.

[18] Lee TD, Genovese ED: Distribution of practice in motor skill acquisition: learning and performance effects reconsidered. Res Quart for Exerc and Sports 1988, 60:59-65.

[19] Maschuw K, Schlosser K, Kupietz E, Slater EP, Weyers P, Hassan I: Do soft skills predict surgical performance?: a single-center randomized controlled trial evaluating predictors of skill acquisition in virtual reality laparoscopy. World J Surg 2011, 35:480-486.

[20] MacRae H, Regehr G, Leadbetter W, Reznick RK: A comprehensive examination for senior surgical residents. Am J Surg 2000, 179:190-193.

[21] Alvand A, Logishetty K, Middleton R, Khan T, Jackson WF, Price AJ, Rees JL: Validating a global rating scale to monitor individual resident learning curves during arthroscopic knee meniscal repair Arthroscopy 2013, 29:906-912.

[22] Hodgins JL, Veillette C, Biau D, Sonnadara R: The knee arthroscopy learning curve: quantitative assessment of surgical skills. Arthroscopy 2014, 30:613-621.

[23] Insel A, Carofino B, Leger R, Arciero R, Mazzocca AD: The development of an objective model to assess arthroscopic performance. J Bone Joint Surg Am 2009, 91:2287-2295.

[24] Bann S, Kwok K, Lo CY, Darzi A, Wong J: Objective assessment of technical skills of surgical trainees in Hong Kong. Br J Surg 2003, 90:1294-9.

[25] Bann S, Davis IM, Moorthy K, Munz Y, Hernandez J, Khan M, Datta V, Darzi A: The reliability of multiple objective measures of surgery and the role of human performance Am J Surg 2005, 189:747-752.

[26] Ezra DG, Aggarwal R, Michaelides M, Okhravi N, Verma S, Benjamin L: Skills acquisition and assessment after a microsurgical skills course for ophthalmology residents. Ophthalmology 2009, 116:257-262.

[27] Faulkner H, Regehr G, Martin J, Reznick R: Validation of an objective structured assessment of technical skill for surgical residents. Am J Surg 1996, 71:1363-1365.

[28] Ilgen JS, Ma IW, Hatala R, Cook DA: A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. Med Educ 2015, 49:161-173.

[29] Kim J, Neilipovitz D, Cardinal P, Chiu M: A comparison of global rating scale and checklist scores in the validation of an evaluation tool to assess performance in the resuscitation of critically ill patients during simulated emergencies (abbreviated as "CRM simulator study IB"). Simul Healthc 2009, 4:6-16.

[30] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner CD, Garrett WE, Olson T, Warme WJ, Nicandri GT: The Arthroscopic Surgical Skill Evaluation Tool (ASSET). Am J Sports Med 2013, 41:1229-1237.

[31] Middleton RM, Baldwin MJ, Akhtar K, Alvand A, Rees JL: Which Global Rating Scale? A Comparison of the ASSET, BAKSSS, and IGARS for the Assessment of Simulated Arthroscopic Skills. J Bone Joint Surg Am 2016, 98:75-81.

[32] Munz Y, Moorthy K, Bann S, Shah J, Ivanova S, Darzi SA: Ceiling effect in technical skills of surgical residents. Am J Surg 2004, 188:294-300.

[33] Alvand A, Auplish S, Gill H, Rees J: Innate arthroscopic skills in medical students and variation in learning curves J Bone Joint Surg Am 2011, 93:e115(1-9).

[34] Grantcharov TP, Bardram L, Funch-Jensen P, Rosenberg J: Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc 2003, 17(7):1082-1085.

[35] Jentzsch T, Rahm S, Seifert B, Farei-Campagna J, Werner CM, Bouaicha S: Correlation Between Arthroscopy Simulator and Video Game Performance: A Cross-Sectional Study of 30 Volunteers Comparing 2- and 3-Dimensional Video Games. Arthroscopy 2016. Epub ahed of print

[36] Rosenthal R, Geuss S, Dell-Kuster S, Schafer J, Hahnloser D, Demartines N: Video gaming in children improves performance on a virtual reality trainer but does not yet make a laparoscopic surgeon. Surg Innov 2011, 2:160-170.

[37] Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ: Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med 2007, 35:883-8.

[38] Harper JD, Kaiser S, Ebrahimi K, Lamberton GR, Hadley HR, Ruckle HC, Baldwin DD: Prior video game exposure does not enhance robotic surgical performance. J Endourol 2007, 21:1207-1210.

[39] Rosser JC, Lynch PJ, Cuddihy L, Gentile DA, Klonsky J, Merrell R: The impact of video games on training surgeons in the 21st century. Arch Surg 2007, 1:181-186.

[40] Thorson CM, Kelly JP, Forse RA, Turaga KK: Can we continue to ignore gender differences in performance on simulation trainers? J Laparoendosc Adv Surg Tech A 2011, 21:329-333.

[41] Konge L, Ali A, Sorensen M, Bitsch M: Gender differences among medical students in the approach to simulation. Ugeskr Laeger 2011, 173:3170-3173.

[42] Elneel FH, Carter F, Tang B, Cuschieri A: Extent of innate dexterity and ambidexterity across handedness and gender: Implications for training in laparoscopic surgery. Surg Endosc 2008, 22:31-37.

[43] Madan AK, Harper JL, Frantzides CT, Tichansky DS: Non-surgical skills do not predict baseline scores in inanimate box or virtual-reality trainers. Surg Endosc 2008, 22:1686-1689.

[44] Doll BB, Jacobs WJ, Sanfey AG, Frank MJ: Instructional control of reinforcement learning: a behavioral and neurocomputational investigation. Brain Res 2009, 1299:74-94.

[45] Strandbygaard J, Bjerrum F, Maagaard M, Winkel P, Larsen CR: Instructor feedback versus no instructor feedback on performance in a laparoscopic virtual reality simulator: a randomized trial. Ann Surg 2013, 257:839-844.

[46] Kahol K, Satava RM, Ferrara J, Smith ML: Effect of short-term pretrial practice on surgical proficiency in simulated environments: a randomized trial of the ‘‘preoperative warm-up’’ effect. J Am Coll Surg 2009, 208:255-268.

[47] Paige JT, Yang T, Suleman R, Chauvin S, Alleyn J, Brewer M, Hoxsey R: Role of instruction method in novices’ acquisition of minimally invasive surgical basic skills. J Laparoendosc Adv Surg Tech A 2011, 21:711-715.

[48] Pernar LI, Smink DS, Hicks G, Peyre SE: Residents can successfully teach basic surgical skills in the simulation center. J Surg Educ 2012, 69:617-22.

[49] Weston MK, Stephens JH, Schafer A, Hewett PJ: Warm-up before laparoscopic surgery is not essential. ANZ J Surg 2014, 84:143-147.

[50] Cowan JB, Seeley MA, Irwin TA, Caird MS: Computer-Simulated Arthroscopic Knee Surgery: Effects of Distraction on Resident Performance. Orthopedics 2016, 139:e240-245.

[51] Kocher H, Warwick J, Al-Ghnaniem R, Patel A: Surgical dexterity after a ‘night out on the town’. ANZ J Surg 2006, 76:110-112.

[52] Bayona S, Akhtar K, Gupte C, Emery RJ, Dodds AL, Bello F: Assessing performance in shoulder arthroscopy: The Imperial Global Arthroscopy Rating Scale (IGARS). J Bone Joint Surg Am 2014, 96:e112(1-7).

APPENDICES

Summary

The validity of arthroscopic simulators and performance tools

As there is a growing demand for more time-efficient and effective methods for medical training without putting patients at risk, the role of simulation keeps expanding. In Chapter 3 it was stated that validation of simulators should precede implementation in medical curricula. However, of the wide variety of commercially available medical simulators, only a small minority was found to be validated. Moreover, substantial heterogeneity and methodological flaws exists among validation studies. This might bring the risk of improper skills training and prevents simulators from becoming established as standard training modalities. Researchers, manufacturers and medical trainers are therefore encouraged to collaborate in order to develop simulators, conduct proper validation studies and select the appropriate simulator for safe and adequate training.

Especially in the field of arthroscopy, a lack of (properly validated) simulators exists. Therefore, the use of simulators for arthroscopic training lag behind compared to other medical areas. In Chapter 4 and 5, two high-fidelity arthroscopic knee simulators were studied for face and construct validity. By using a standardized protocol and validated outcome measures, several important limitations have been overcome, and basic level of validity were demonstrated. Moreover, the studies revealed both simulators require further improvement. There is an absence of transfer validity evidence of arthroscopic knee simulators, due to a lack of randomized controlled trials covering this level of validity. In Chapter 6, a pilot study was conducted to develop a solid protocol for a large scale randomized controlled transfer validity study. Using a simulator-trained group and a theoretically trained control group, appropriateness and feasibility of the used methods and procedures were investigated. Recommendations were made for appropriate study period, assessment tools and methods, and outcome measures, to enable an RCT

that will offer clarity about the effectiveness and transfer validity of arthroscopic knee simulators.

The acquirement of arthroscopic skills is stimulated by the use of objective tools and metrics to monitor arthroscopic performance and provide feedback. As metrics and complementary thresholds for safety performance during arthroscopy are scarcely studied, Chapter 7 elaborated on the maximum allowed forces for safe joint stressing. It was stated that thresholds for safety metrics are necessary for efficient training, and that force feedback with safety levels should be implemented in arthroscopic knee simulators in order to give orthopaedic residents the opportunity to practice safe tissue manipulation. Besides task-specific metrics, more holistic tools for performance tracking are required as well. In Chapter 8, construct validity of two Global Rating Scales, designed for evaluation of arthroscopic performance was studied. Although it was suggested repeatedly that GRSs are suitable to establish individual learning curves, the results of this study demonstrated that these scales are not sufficiently sensitive and consistent to establish individual learning curves. They are however suitable to objectively evaluate global progress of residents in the operating room when acquiring arthroscopic skills, in particular on group level.

In recent years, the number of (high-fidelity) arthroscopic simulators and performance tracking tools have increased. However, for both simulators and assessment tools, more evidence on their validity and usefulness is needed. Of particular importance for the acceptance of simulators as a useful tool in arthroscopic education is the demonstration of transfer validity. Future studies should focus on high-quality study designs and standardized protocols in order to offer orthopedic residents an adequate training environment to acquire and maintain arthroscopic

skills, and to enable performance tracking and proficiency assessment, with the ultimate goal to improve and secure patient safety.

Samenvatting

De validiteit van arthroscopische simulatoren en prestatie parameters

Aangezien er een groeiende vraag is naar meer tijdbesparende, effectieve en patiënt-veilige trainingsmethoden binnen medische curricula, wordt simulatie steeds belangrijker. In Hoofdstuk 3 werd gesteld dat simulatoren gevalideerd moeten worden voordat ze worden geïmplementeerd in medische curricula. Van de grote verscheidenheid aan commercieel beschikbare medische simulatoren, is echter slechts een kleine minderheid gevalideerd. Bovendien zijn de bestaande validatie studies zeer heterogeen en is er sprake van een scala aan methodologische tekortkomingen. Dit vergroot het risico op het incorrecte vaardigheidstraining, en werkt de in gebruik name van simulatoren als standaard training module tegen. Onderzoekers, fabrikanten en opleiders worden daarom aangemoedigd samen te werken om simulatoren te ontwikkelen, gedegen validatie studies uit te voeren en de juiste simulator voor een veilige en adequate training te selecteren.

Binnen de arthroscopische chirurgie, is er een gebrek aan (goed gevalideerde) simulatoren. In Hoofdstuk 4 en 5 zijn twee *high-fidelity* arthroscopische knie simulatoren getest op face en construct validiteit. Door gebruik te maken van een gestandaardiseerd protocol en gevalideerde uitkomstmaten zijn een aantal belangrijke beperkingen ondervangen, en werden basis levels van validiteit aangetoond. Bovendien bleek dat bij beide simulators verdere verbetering vereist is. Er is nog altijd geen bewijs van transfer validiteit van artroscopische knie simulatoren, vanwege een gebrek aan gerandomiseerde studies die dit validiteitsniveau bestuderen. In Hoofdstuk 6 is een pilot studie uitgevoerd om een solide protocol voor een grootschalige gerandomiseerde studie te ontwikkelen, om transfer validiteit te kunnen bestuderen. Door een simulator-getrainde groep en een theoretisch getrainde controlegroep met elkaar te vergelijken, is de geschiktheid en haalbaarheid van gebruikte

methoden en procedures onderzocht. Vervolgens zijn er aanbevelingen gedaan voor het realiseren van een afdoende studie periode, evaluatietesten, methoden en uitkomstmaten, die zouden moeten leiden tot een RCT die duidelijkheid kan geven over de effectiviteit en de transfer validiteit van arthroscopische knie simulators.

Het aanleren van arthroscopische vaardigheden wordt gestimuleerd door objectieve parameters en meeteenheden om arthroscopische competentie te monitoren en feedback te geven. Aangezien meetbare veiligheidswaarden en de daarbij horende drempelwaarden tijdens artroscopieën nog nauwelijks bestudeerd zijn, zijn in Hoofdstuk 7 de maximaal toegestane krachten op knieligamenten onderzocht. We stelden dat arthroscopische training drempelwaarden voor meetbare veiligheidswaarden behoeft, en zouden moeten worden geïmplementeerd in arthroscopische knie simulatoren om orthopeden in opleiding de mogelijkheid te bieden zich het veilig omgaan met het knieweefsel eigen te maken. Naast taak-specifieke meetwaarden, is er ook behoefte is aan meer holistische instrumenten voor het beoordelen van prestaties. In Hoofdstuk 8 is construct validiteit van twee Global Rating Scales, specifiek ontwikkeld voor het evalueren van arthroscopische vaardigheden, bestudeerd. Hoewel herhaaldelijk is gesuggereerd dat GRSsen geschikt zijn voor het volgen van individuele leercurves, hebben de resultaten van deze studie getoond dat deze schalen onvoldoende gevoelig en consistent zijn om individuele leercurve te bewerkstelligen. Zij zijn echter wel geschikt om globale vooruitgang van arthroscopische vaardigheden in de operatiekamer te objectief beoordelen, met name op groepsniveau.

De afgelopen jaren is het aantal (*high-fidelity*) arthroscopische simulators en parameters om prestaties mee te monitoren toegenomen.

Echter, voor zowel simulatoren als beoordelingsmethoden geldt dat meer bewijs van hun validiteit en bruikbaarheid nodig is. Van bijzonder belang voor de acceptatie van simulatoren als arthroscopische trainingsmodaliteit is het bewijs van transfer validiteit. Toekomstig onderzoek moet zich richten op het ontwikkelen van hoogwaardige studies en gestandaardiseerde onderzoeksprotocollen. Zodoende kan orthopeden in opleiding een adequate training worden geboden om arthroscopische vaardigheden te verwerven en te onderhouden, en om vaardigheid en vooruitgang te evalueren, met het uiteindelijke doel de veiligheid van de patiënt te bevorderen en te waarborgen.

List of contributing authors

Jenny Dankelman, PhD
Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands
Supervision

Niek C.N. van Dijk, MD, PhD
Department of Orthopedic Surgery, Academic Medical Center, Amsterdam, The Netherlands
Supervision

Jason Dragoo, MD, PhD
Department of Orthopedic Surgery Stanford University School of Medicine, Standford, United States
Supervision and data collection support

Gino M.M.J. Kerkhoffs, MD, PhD
Department of Orthopedic Surgery, Academic Medical Center, Amsterdam, The Netherlands
Study design, supervision and data collection support

Mario Maas, MD, PhD
Department of Radiology, Academic Medical Center, Amsterdam, The Netherlands
Supervision

Bas van Ooij, MD
Department of Orthopedic Surgery, Academic Medical Center, Amsterdam, The Netherlands
Data collection support

Thijs Pahlplatz, Msc
Department of Orthopedic Surgery, Academic Medical Center, Amsterdam, The Netherlands
Data collection support

Matthias U. Schafroth, MD, PhD
Department of Orthopedic Surgery, Academic Medical Center, Amsterdam, The Netherlands
Data collection support

Drs. Inger N. Sierevelt
Spaarne Hospital of Hoofddorp, the Netherlands
Methodological and statistical support

Tim Horeman, PhD
Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands
Data collection support

Gabrielle J.M. Tuijthof, PhD
Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands
Study design, supervision and data collection support

Pieter-Bas H.L.M. Wulms, Msc
Department of Biomechanical Engineering, Delft University of Technology, Delft, The Netherlands
Data collection support

Portfolio

Name PhD student:	Jonáh Juana Stunt
PhD period:	September 2012-July 2016
Name PhD supervisor:	Prof. dr G.M.M.J Kerkhofs, Prof. dr. C.N. van Dijk

1. PhD training

	Year	Work-load (ECTS)
General courses		
- Practical Biostatistics	2013	1.1
- Clinical Data Management	2015	0.3
Specific courses		
- Advanced Biostatistics in R	2015	2.1
- Computing in R	2015	0.4
- Data Analysis in Matlab	2015	0.7

Presentations

-	J.J. Stunt, P.H.L.M. Wulms, G.M.M.J. Kerkhoffs, I.N. Sierevelt, M.U. Schafroth, G.J.M. Tuijthof, 'Variation in joint stressing magnitudes during knee arthroscopies' DSSH Congres, Amsterdam	2013	0.5
-	J.J. Stunt, G.M.M.J. Kerkhoffs, T. Pahlplatz, M. Maas, G.J.M. Tuijthof, 'Transfer validity of simulator training for knee arthroscopy: a randomized controlled pilot study', DSSH Congres, Delft	2016	0.5

Other

-	Rapid Miner	2016	-
-	Data Science en Big Data Analytics	2015-2016	-
-	SQL training	2015	-
-	Journal Club	2012-2013	2

2. Parameters of Esteem

		Year
Grants		
-	European Society of Sports Traumatology Knee Surgery and Arthroscopy (ESSKA): €10.000	2013
-	Marti Keuning Eckhart stichting, Lunteren, the Netherlands: €17.000	2012

3. Publications

Peer reviewed	**Year**
- J.J. Stunt , P.H.L.M. Wulms, G.M.M.J. Kerkhoffs, prof. J. Dankelman , C.N. van Dijk, G.J.M. Tuijthof, 'How valid are commercially available simulators' *Advances in Medical Education and Practice*	2014
- J.J. Stunt, G.M.M.J. Kerkhoffs, C.N. van Dijk, G.J.M. Tuijthof, 'Validation of the ArthroS virtual reality simulator for arthroscopic skills' *Journal of Knee Surgery, Sport traumatology and arthroscopy*	2014
- J.J. Stunt, G.M.M.J. Kerkhoffs, T. Horemans, C.N. van Dijk, G.J.M. Tuijthof , 'Validation of the PASSPORT V2 training environment for arthroscopic skills' *Journal of Knee Surgery, Sport traumatology and arthroscopy*	2014
- J.J. Stunt, P.H.L.M. Wulms, G.M.M.J. Kerkhoffs, I.N. Sierevelt, M.U. Schafroth, G.J.M. Tuijthof , 'Variation in joint stressing magnitudes during knee arthroscopies' *Journal of Knee Surgery, Sport traumatology and arthroscopy*	2013
- J.J. Stunt, B. van Ooij, G.M.M.J. Kerkhoffs, I.N. Sierevelt, M.U. Schafroth, C.N. van Dijk, J. Dragoo, G.J.M. Tuijthof, 'Global rating scales to objectify arthroscopic performance'	submitted
- J J. Stunt, G.M.M.J. Kerkhoffs, T. Pahlplatz, M. Maas, G.J.M. Tuijthof Transfer validity of simulator training for knee arthroscopy: a randomized controlled pilot study.	submitted

Dankwoord

Zonder de hulp van vele mensen om mij heen was dit proefschrift niet tot stand gekomen. Allereerst veel dank aan alle patiënten, arts-assistenten, stafleden en collega's voor het meewerken aan de verschillende studies. Tevens veel dank aan de derdejaars Geneeskunde studenten die hebben deelgenomen aan de randomized controlled trial. Jullie waren een enthousiaste, leergierige en gezellige groep, en het was me een genoegen mijn onderzoeksperiode met jullie tot een einde te brengen.
In het bijzonder gaat mijn dank uit naar:

Promotor prof. dr. C.N. van Dijk. Heel veel dank voor het bieden van de mogelijkheid te promoveren op uw afdeling.

Promotor prof. dr. G.M.M.J. Kerkhoffs. Gino, ongelooflijk veel dank voor alles. Dank voor het meedenken, voor je hulp bij alle onderzoeken, voor je kritiek, je expertise, voor de leuke gesprekken, je begrip, morele steun en peptalks. Wanneer de moed mij in de schoenen zonk en ik gestrest en in paniek naar je toe kwam, kon ik mijn hart bij je ophalen. Het was zo fijn voor me dat jij de boel weer een beetje kon relativeren en me het gevoel gaf dat het allemaal wel goed kwam. Ik ben je ontzettend dankbaar.

Co-promotor dr. ir. G.J.M. Tuijthof. Gabrielle, ik kan je moeilijk uitleggen hoe groot mijn dank aan jou is. Jouw bijdrage aan dit proefschrift is van onschatbare waarde. Het botste weleens tussen ons en we waren het niet altijd met elkaar eens. Maar dat geeft niet, sterker nog, het was -in elk geval voor mij- ook wel goed. Je bent een fantastische leermeester ge-

weest: veeleisend en kritisch, doortastend en recht door zee, intelligent en met een uitzonderlijke energie. Dat werd al duidelijk in een van je eerste mails aan mij, waarin je schreef: ‘en ja, ik mik op een hele stijle leercurve van jou, maar geef ook aan als het te hard gaat, oke?’. Je arbeidsethos en kritische houding waren koren op mijn molen. Ik heb ontzettend veel respect voor je, als mens en als wetenschapper. Ik heb met veel bewondering gekeken naar hoe jij je staande weet te houden in de weerbarstige omgeving waarin jij je begeeft. Je bent me ontzettend veel geleerd; over wetenschappelijk onderzoek doen en over wat het betekent wetenschappelijk onderzoeker te zijn, maar ook over mezelf. Ongelooflijk veel dank voor alle kennis en inzichten, en de tijd, moeite en energie die je in mij en mijn onderzoek hebt gestoken.

Secretariaat G4. Rosalie, bijzondere dank voor alle hulp bij de afronding van mijn promotietraject en je oog voor de details.

Dr. ir. Leendert Blankevoort, dank voor het meedenken en voor de adviezen, maar vooral voor de wetenschappelijke houding en blik die je ons tijdens de Journal Club uurtjes probeerde bij te brengen.

Alle co-auteurs, prof. dr. Jenny Dankelman, prof. dr. Niek C.N. van Dijk, dr. Jason Dragoo, dr. ir. Tim Horeman, prof. dr. Gino M.M.J. Kerkhoffs, prof. dr. Mario Maas, drs. Bas van Ooij, Thijs Pahlpatz Msc, dr. Matthias U. Schafroth, drs. Inger N. Sierevelt, dr. ir. Gabriëlle J.M. Tuijthof en P.H.L.M. Wulms Msc. Veel dank voor jullie respectievelijke bijdragen en kritische blik.

Overige leden van de promotiecommissie, prof. dr. M. Maas, prof. dr. M.P. Schijven, prof. dr. J. Dankelman, prof. dr. R.L. Diercks, prof. dr. M.J.H.M. Lombaerts en dr. M.C. Obdeijn. Heel veel dank voor het beoordelen van mijn proefschrift en voor de bereidheid de rol van opponent aan te nemen.

Goos Huijzer, Peter Schenk, Joeri Post, André Ermers, Casper Scheuer, Willem Nerkens, Monica Moreo, Kim Opdam, Thijs Pahlpatz, Gwen Vuurberg en Yaël Ashruf voor jullie onmisbare hulp bij de uitvoering van de randomized controlled trial. Zonder jullie was de stress me teveel geworden.

Mijn roomie Ruben, en natuurlijk Inge, Arthur, Wout, Rogier, Hanneke en Joy: jullie aanwezigheid, gezelligheid en flauwe grappen hebben mijn jaar bij de orthopedie kleurrijk en vrolijk gemaakt. Ondanks het feit dat ik een buitenbeentje en vreemde eend in de bijt was, heb ik me bij jullie gewaardeerd en op mijn gemak gevoeld en koester ik mooie herinneringen aan de congresbezoeken.

Mijn R&D-kamergenoten en collega's van iCOV voor hun steun, interesse en zorg. Gees en Curd, jullie in het bijzonder, omdat jullie mij de gelegenheid gaven om de eindsprint te trekken ('volgende week ben je 'ziek' en blijf je dus thuis'), maar bovenal voor jullie lieve en dierbare steun, medeleven en begrip.

Cees Lucas, omdat je een belangrijke rol hebt gespeeld in verschillende van mijn overwegingen. Veel dank voor je luisterend oor, wijze adviezen, en het door jou geschreven treffende vers dat je me te lezen gaf.

Inger, heel veel dank! Niet alleen voor je hulp bij de methodologische aspecten van mijn onderzoeken en het sparren over statistische analyses, maar ook voor de leuke gesprekken over sport, muziek en je kinderen, voor je betrokkenheid en voor je morele steun.

Jair en Maxim. Ofschoon jullie niet gehinderd zijn door enige kennis met betrekking tot het onderwerp van mijn proefschrift, en in geval van een black-out ik inhoudelijk helemaal niets aan jullie heb, voel ik me met mijn lieve grote broertjes aan mijn flanken gesterkt en minder bang de oppositie te voeren. Bovendien weet ik zeker denk ik dat ik de leukste en knapste paranimfen *in history* heb. Ik ben ontzettend trots en dankbaar dat jullie mijn paranimfen willen zijn.

Al mijn lieve vrienden en familieleden, omdat jullie met mij hebben meegeleefd en zo begaan met mij waren. Lieve Dunja, ontzettend bedankt voor het eindeloos aanhoren van mijn gedub en getwijfel, en voor je hulp bij het nemen van belangrijke beslissingen. Bijzondere dank aan Felix, voor het verdragen van mijn gestress en frustraties en gepraat over simulator studies, en voor het verlenen van een schouder om op te huilen als dat nodig was. Lieve Jair, ik ben je zeer erkentelijk voor de energie die jij me gaf voor de laatste loodjes. Liefste Jaja, ik vind het zo ontzettend bijzonder en dierbaar dat jij de omslag van dit boekje hebt ontworpen en gemaakt, je weet niet half wat het voor me betekent. Grazie! Gideon, mijn liefste en beste vriend, zoveel dank, omdat jij mijn steun en toeverlaat bent (geweest). Tot slot, lieve lieve papa en mama, oneindig veel dank, omdat zonder jullie, tja, wat wel…?

Curriculum vitae

Jonáh Juana Stunt was born in Amsterdam on the 23th of May, 1985. After graduating from high school (Gymnasium Haganum, the Hague) in 2003, she followed several courses to qualify for Medical School. She started studying Medicine at the University of Amsterdam in 2004, and in 2005 she started studying Cultural Anthropology and Sociology of non-Western Societies as well. To reconsider whether Medical School was the right choice, she took a one-year sabbatical to travel through Southeast Asia, to deliberate and cogitate. Because of her affection for research, science and the human brain, she continued Medical School and completed a so-called Free Doctorate, specializing in Neuroscience. At the same time she continued studying Cultural Anthropology and Sociology of non-Western Societies. She fulfilled her medical study with a one-year research internship at the Dutch Institute for Neuroscience (NIN), working on research on Alzheimer's disease and Multiple Sclerosis. Following a short research internship under the supervision of Gabrielle Tuijthof, she started working as a PhD student at the Department of Orthopedic Surgery of the Academic Medical Centre in September 2012. She performed the larger part of her research as an external PhD candidate, while doing her masters in Evidence-Based Practice, to further develop her scientific and methodological knowledge and skills. In addition, she started working as a scientific researcher at iCOV, a governmental organization that investigates financial crime. She currently works as coordinator of the Research & Development department at iCOV.