Linköping University medical dissertations
No. 1167

Construction, Validation and Application of a Virtual Reality Simulator for the Training of Transurethral Resection of the Prostate

Reidar Källström

Linköping University
FACULTY OF HEALTH SCIENCES

Division of Surgery and Division of Urology
Department of Clinical and Experimental Medicine
Faculty of Health Science, Linköping University
SE-58185 Linköping, Sweden

Linköping 2010

Front cover is designed by Lina Källström

Printed in Sweden by LiU-Tryck, Linköping, 2010

ISBN 978-91-7393-444-2
ISSN 0345-0082

If you put tomfoolery into a computer, nothing comes out of it but tomfoolery. But this tomfoolery, having passed through a very expensive machine, is somehow ennobled and no-one dares criticize it.

Pierre Gallois

Contents

Abstract

The fundamental of surgical training is the traditional apprenticeship method introduced by William Halsted which has been used for the last 100 years. It is based on learning in the operating room (OR) where the resident is guided by an experienced surgeon and gradually and methodically exposed to surgery. The continuous development of surgical methods together with the growing awareness of medical errors and ethical considerations have made the Halsted method outdated and there is an obvious need to be able to learn the skills of surgery without risking patient safety. New methods such as laparoscopy and endoscopy demand specific skills and abilities that may not be met by everyone. At the same time, the physical limitations of these new methods have made it possible to construct virtual reality (VR) simulators to practise and learn the skills necessary.

This study is about the construction and evaluation of a VR-simulator for the training of transurethral resection of the prostate (TURP). It also concerns the specific abilities needed to become a good surgeon.

A simulator for training TURP was developed after a face validity study where 17 experienced urologists gave their opinion of the specific content necessary for the training of this procedure. After a content validity study by nine experienced urologists and application of necessary improvements, a group of 11 medical students and nine experienced urologists performed a construct validity test where the urologists showed significantly higher levels of both skill and effectiveness compared to the inexperienced students when performing a simulated TURP procedure. The students showed a positive learning curve, but did not reach the levels of the urologists. The results of the experienced urologists were used as the minimal criterion level when 24 urology residents practised the procedure. Training took place while on a course on benign enlargement of the prostate and its treatment options, with emphasis on the “gold standard” treatment – TURP. During the course they performed three guided and video-taped TURP-procedures each on selected patients. Between two of the procedures they performed criterion-based training in the simulator. This VR-to-OR study showed improvement in operative skills with the same patient outcome as in the normal clinical situation. It also showed that simulator training improved their skills even more. During their time on the course their personality traits (TCI) and cognitive abilities (Rey complex figure and recognition trial, tower of London, WAIS-III) were tested. The results showed that a better learning curve in the OR was associated with a better simulator learning curve and a good visuospatial memory. The associated personality traits were high levels of goal directedness, impulse control, responsibility, anticipation of harmful events and a balanced attachment style.

In conclusion, we have demonstrated that it was technically possible to construct a useful simulator for the training of TURP (PelvicVision®) which may now be considered clinically validated for this purpose. Novice training and performance in the simulator improves the learning curve and predicts the resident’s performance in the OR. The results support the implementation of validated simulation technology in a criterion-based training curriculum for residents. Furthermore, the results showed preliminary data on personality traits and visuospatial abilities that are important for learning a complex surgical procedure.

Key words*: surgical education, simulation, transurethral resection of prostate, psychometric tests, personality, validation, virtual reality, proficiency based training, objective assessment*

List of original papers

This thesis is based on the following papers, which will be referred to by their roman numerals:

I. Källström R, Hjertberg H, Kjölhede H, Svanvik J.

Use of a virtual reality, real-time, simulation model for the training of urologists in transurethral resection of the prostate.

Scand J Urol Nephrol 2005; 39: 313-320.

II. Källström R, Hjertberg H, Svanvik J.

Construct Validity of a Full Procedure, Virtual Reality, Real-Time, Simulation Model for Training in Transurethral Resection of the Prostate.

J Endourol 2010; 24:109-115.

III. Källström R, Hjertberg H, Svanvik J.

Impact of VR-simulated training on urology residents' performance of transurethral resection of the prostate.

Submitted 2009, J. Endourol.

IV. Källström R, Rousseau A, Bengtsson A, Hjertberg H, Svanvik J.

Simulator performance, psychometrics and personality testing guiding the choice of clinical discipline.

Manuscript

The published papers are reproduced with the permission of Taylor & Francis (Scandinavian Journal of Urology and Nephrology) and Mary Ann Liebert Inc (Journal of Endourology)

Abbreviations

BPH	Benign Prostatic Hyperplasia
CT	Computed Tomography
CUSUM	Cumulative Summation Analysis
DSB	Digit Sequence Backward
DSF	Digit Sequence Forward
EAES	European Association for Endoscopic Surgery
FDA	Food and Drug Administration
IPSS	International Prostate Symptom Score
ITER	In-Training Evaluation Report
LNS	Letter-Number Sequencing
LUTS	Lower Urinary Tract Symptoms
OR	Operating Room
OSATS	Objective Structured Assessment of Technical Skills
OSCE	Objective Structured Clinical Examination
RCFRT	Rey Complex Figure and Recognition Trial
RCT	Randomised Controlled Trial
SPSS	Statistical Package for the Social Sciences
TCI-R	Temperament and Character Inventory - Revised
ToLdx	Tower of London, Drexel University
TRUS	Transrectal Ultrasonography
TURB	Transurethral Resection of Bladder Tumours
TURP	Transurethral Resection of the Prostate
VIST	Vascular Intervention Simulation Trainer
VR	Virtual Reality
WAIS	Wechsler Adult Intelligence Scale

Introduction

The cornerstone of surgical training programme is the traditional apprenticeship model. This model consists of three key components: observation, coaching and training. Sir William Halsted (1852-1922)[1] is said to be the father of this method in surgical training. He introduced the apprenticeship model of graded responsibility in USA, as an adaptation of the German residency training programme. The Halsted method of learning is based on the methodical exposure to clinical experience in the operating room (OR), under the close guidance of dedicated senior attending surgeons, during several years of residency. In this discussion he represents the idealistic concept of the brave, skilful and innovative surgeon. Halsted was said to be a poor student who never checked out a book from the library at Yale College, but excelled in medical school at Columbia University College where he graduated 1877, near the top of his class. He travelled to observe and learn from surgeons and scientists in Europe and went back to USA in 1880. By then he was characterised as a bold, daring, original and indefatigable surgeon. He performed, for instance, one of the first gallbladder procedures in the USA and one of the first blood transfusions. He also proved that injection of cocaine into a nerve can give effective local anaesthesia. Halsted moved to Baltimore to join the staff at Johns Hopkins Hospital where he was reputed to be a slow, methodical and careful surgeon. It was at Johns Hopkins he started the first formal surgical residency training programme in the USA. The programme consisted of an internship of undefined length, the individuals advanced when Halsted believed they were ready for the next level. As the father of “safe” surgery, he promoted state of the art surgical principles: control of bleeding, accurate anatomical dissection, exact approximation of tissue in wound closure without excessive tightness and gentle handling of tissues. He also used the principle of complete sterility and invented the surgical glove.

Much has changed since Halsted’s days. The demand for precision is rising and affects all aspects of surgery, from diagnosis and selection of treatment, to procedure time, turnover and health-care time. The constant development of surgical techniques and medical knowledge makes the “old time” surgeon, who performed all kinds of surgical procedures, obsolete. To maintain high quality performance and to be updated on all current knowledge is hard work even for a highly specialised surgeon today. This is reflected in the current training and education of surgical residents. In the teaching hospitals there is a high turnover of patients, who often have complicated conditions. The growing awareness of medical errors, high result expectancy and the ethical aspects of training on patients limits the possibilities to train surgeons the “Halsted” way. There is a call for training methods that do not put the patient at risk. The introduction of laparoscopic methods in the late 1980s gave rise to a peak in complications and an increased awareness of surgical skills. The laparoscopic technique provides a two-dimensional picture of the operative field which leads to problems in eye-hand coordination and cognitive mapping. Long instruments are used through the abdominal wall and movements of the surgeon's hand is inversed by the fulcrum effect[2]. This technique put new demands on visuospatial and psychomotor abilities which not all surgeons master. The same limitations, however, made it possible to create computer-based virtual reality simulators to practise these skills. The first commercially available VR-simulators were introduced in the late 1990s. At the same time a prominent report "To Err is Human" was published where it was estimated that 44-98 000 patients die every year in the USA due to medical errors. About 80% occurred in hospitals and it was estimated that about 50% of the errors were preventable. Most of the errors were due to drug complications, but a large number were related to surgery. The development of new techniques and the increasing awareness of preventable errors put high demands on ongoing training of the medical staff.

Since the beginning of this century the development and evaluation of the use of medical simulators has avalanched and there are today a number of validated VR-simulators for the training of various medical procedures. The need for well-designed simulators for specific procedures is still great and it is important that new teaching aids have been shown to have a positive learning effect that is transferable to the real procedure before they are introduced into medical teaching programmes. This thesis concerns the research and development of such teaching aids. The change in attitude from the brave and skilled to the careful and highly specialised surgeon of today may also be important in the learning situation and the choice of career. What is the characteristic personality of a surgeon and are there personal traits that are favourable for modern surgical skills?

Transurethral resection of the prostate

Transurethral resection of the prostate (TURP) is the gold standard for treatment of benign prostatic hyperplasia (BPH)[3,4]. BPH is a pathologic process that contributes to lower urinary tract symptoms (LUTS) in aging men. The underlying aetiology of prostatic growth has not been established. Androgens are a necessary but not a clearly causative aspect of BPH. LUTS are not only due to a mass-related increase in urethral resistance, a significant portion is due to age-related detrusor dysfunction. Bladder outlet obstruction itself may induce a

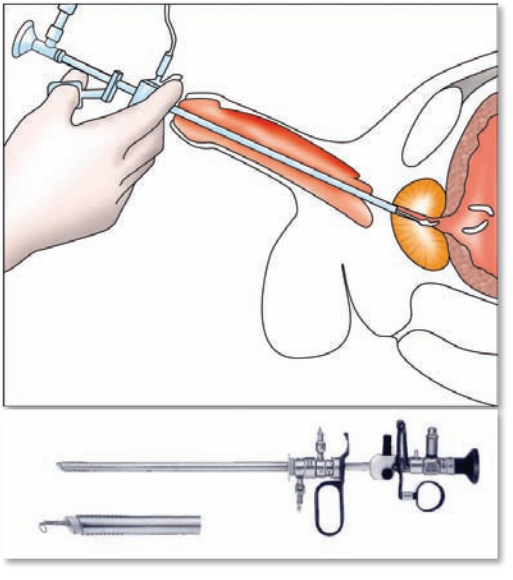

Figure 1. A Resectoscope. Used for transurethral resections.

variety of neural alterations in the bladder, which contribute to the symptomatology. The most common reasons for recommending intervention in a patient with symptoms of bladder outlet obstruction and irritability are that the symptoms interfere with the patient's quality of life. Although symptoms constitute the primary reason for recommending intervention, in patients with an obstructing prostate there are some absolute indications (acute urinary retention, recurrent infection, recurrent haematuria, and azotemia)[5].

When performing the TURP-procedure, the patient is placed in the lithotomy position, and usually under spinal anaesthesia. The resection is performed using a resectoscope (Figure 1) with current applied to the wolfram electrode at the tip of the instrument. The current can be alternated between cutting and coagulation. There is also a video camera connected to the optics to show the resectoscope view on a video screen. Various surgical techniques have been espoused by urologists for removing the prostate adenoma. The resection technique may vary but should be based on an orderly plan in a step-by-step manner. The method used in this study is a modification of the method described by Nesbit in 1943. The procedure starts with the resection of the median lobe as far distal as the ***verumontanum*** (Figure 2). The sphincter mechanism is located distal to the level of the verumontanum and one main principle is to never resect any adenoma distal to this level. The resection then continues to make an incision/resection at the one or eleven o'clock position and distally to cause the lateral lobe to fall downwards and also to take care of the main arterial supply of the adenoma. The large lateral lobe can then be resected with limited amount of blood loss. The same procedure is performed on the contralateral side. Finally some trimming is done to get rid of still remaining adenoma, including the area close to the verumontanum. During the entire procedure almost every cut will open a blood vessel and bleeding will impair the vision. Open blood vessels can be

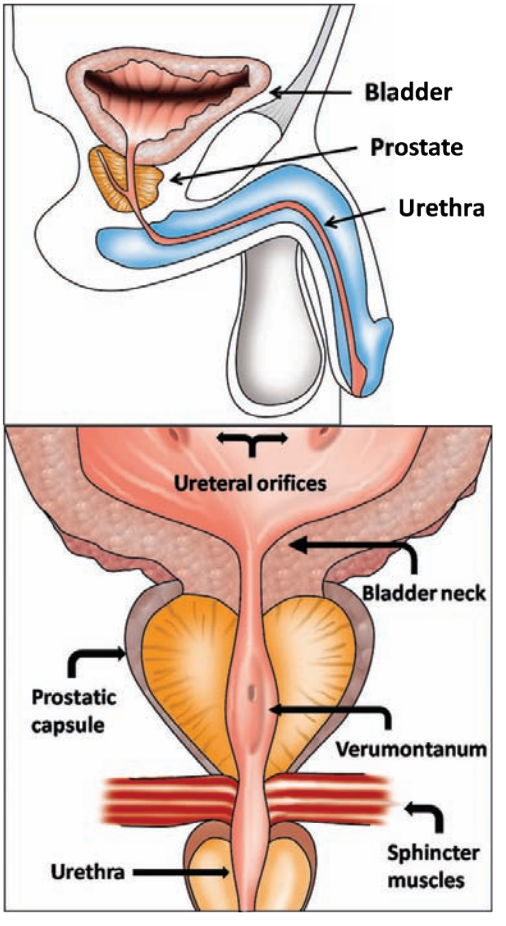

Figure 2. The position and anatomy of the prostate

sealed by applying a coagulating instead of a cutting current as controlled by foot-pedals. Blood and resected pieces of adenoma will obscure the vision unless an irrigation solution is administered via channels in the resectoscope causing a flow of fluid that clears the view. This fluid is stored in the bladder and the flow will gradually diminish as the capacity of the bladder is reached. The fluid, blood and adenoma chips in the bladder are emptied via the resectoscope by removing the optics and cutting instrument or via a troachar through the abdominal wall. There are also techniques for continuous irrigation of fluid via the resectoscope.

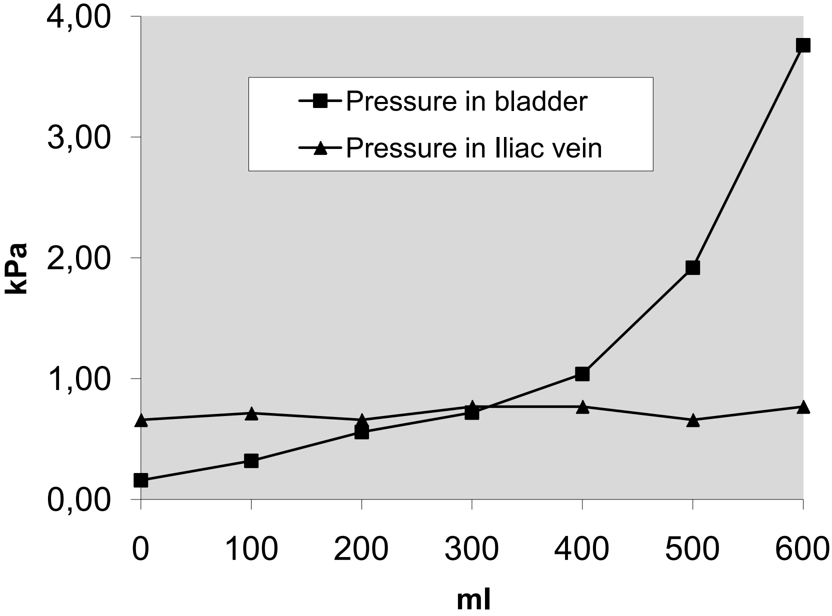

Figure 3. Pressures in the bladder and iliac vein with increasing volume in the bladder according to Hultén et al.[6]

During the procedure it is necessary to obtain a good balance between cutting and coagulating in order to resect the necessary amount of adenoma without draining the patient of too much blood. The frequency of transfusion due to haemorrhage is reported to be between 0.4-7.1% [7]. Intraoperative problems apart from blood loss are TUR-syndrome and extravasation which both are caused by the irrigation solution. The irrigation solution must not contain any electrolytes that would disturb the effect of the cutting/coagulating current and isotonic or hypotonic but non-haemolytic fluids are used. To achieve flow of irrigation fluid a pressure gradient is applied, often by suspending the irrigation fluid bag above the patient. When the bladder is empty the pressure is low but increases when the bladder capacity is reached (Figure 3). If this pressure exceeds that in the blood vessels it may result in systemic uptake of irrigation fluid and give rise to the TUR-syndrome caused by hyponatraemia and/or overhydration. This is a rare (<1%)[7] but serious complication with reported mortality[7-10]. Extravasation of irrigation fluid occurs when there is a perforation of the prostatic capsule. The fluid then accumulates in the surrounding fat causing pain and nausea despite spinal anaesthesia. For a list of possible complications during TURP see Table 1.

The procedure is easy to explain but has a long learning curve. It is reported that it takes up to 80-100 procedures before proficiency is reached[11-13].

Complications when performing TURP	Frequency %	Avoidable cause
Clot retention	1.3 - 11.0	Poor haemostasis control
Bleeding & transfusion	0.4 -22.0	Poor haemostasis control
TUR-syndrome	0.0 - 2.8	Perforation of capsular veins or sinuses
Capsular perforation or bladder neck division with extravasation of irrigation solution	0.9 - 10.0	Poor resection control
Hydronephrosis	0.0 - 0.3	Injury of ureteral orifices
Epididymitis / urinary tract infection	1.6 - 25.0	Long duration of procedure, clot retention
Urosepsis	0.0 - 3.0	Long duration of procedure, clot retention
Failure to void	3.0 - 7.1	Mainly due to primary detrusor failure, may be caused by overextension
Incontinence	0.3 - 38.0	About 0.5% due to trauma of the external sphincter muscle, poor resection control
Urethral stricture	2.2 - 9.8	Inappropriate size or insufficient electrical isolation of the resectoscope
Bladder neck stenosis	0.3 - 9.2	No bladder neck incision made
Retrograde ejaculation	53 - 75	May be avoided by sparing tissue around the verumontanum
Recurrent BPH	0.0 - 6.6	Insufficient resection
Peroperative mortality	0.1 - 0.23	Selection of patients

Table 1. Complications associated with transurethral resection of the prostate[7]

The history of transurethral resection of the prostate

Already 3000 B.C. the Egyptians knew about transurethral entry into the bladder, using instruments made of copper and lead. Hindus used tubes of gold, iron or wood to dilate the urethra and metal catheters and probes were found in the ruins of Pompeii. Around 1800 cutting instruments were developed for use in the surgery of bladder stones. Morgagni made one of the first descriptions (1719) of an enlarged prostate in a patient who died of urinary retention. In the early days there was a considerable amount of complications associated with the TURP procedure. Mortality rate was described to be between 8-50% and urinary leakage due to damages to the urethral sphincter was common. The alternatives to this treatment were open surgery with great risks for complications and postoperative care for several weeks or permanent catheter. Open surgery was the standard treatment in Europe until beginning of the 1960s. Change to the transurethral approach followed the inventions by H.H. Hopkins, professor of optics, who developed light transmitters of glass fibres and rod lenses. During the 1970s inventions have led to diathermy generators using semiconductors and in the late 1980s the computer chip video camera was added[14,15] (Table 2).

1719	Morgagni	Description of an enlarged prostate in a patient who died of urinary retention
1726	La Faye	Creation of a passage through the prostate and bladder neck
1807	Bozzini	Tube and wax candle in a container. Illumination of the inside of the bladder
1834	Guthrie	Folding knife for cutting the bladder neck. Poor results, catheters used instead.
1840	Mercier	Instrument to pinch out tissue from the bladder neck
1853	Desourmeaux	Lens connected to the tube, terpentin-alcohol burner as light source
1861	Tripier	Galvanic current through the prostate via electrodes in urethra and rectum. Probably no effect at all.
1877	Bottini	Galvanic instrument with heated platinum plate at the tip, applied to bladder neck purely by sense. Some of the surviving patients reported improved symptoms
1879	Nitze, Leiter	Glowing thread of platinum at the tip of the instrument as light source
1880	Edison	Electric light
1890	d'Arsonval	High frequency (10 000 Hz) alternating current can pass tissue without muscle contractions

1900	Freudenburg	Added optics
1907	Pozzi	High frequency sparks to treat skin tumours (fulguration; Latin fulgur = lightning)
1909	Beer	Electric cautery underwater
1909	Young	Punch to cut out prostatic tissue. Problem with bleeding
1910	Beer	High frequency sparks to treat bladder tumours
1920	Caulk	Electrically heated blade. Control of bleeding
1923	Wappler	Tube generator creating a sinus-shaped high frequency alternating current to cut tissue
1925	Walker	Instrument isolated with Bakelite, high frequency alternating current to coagulate the tissue before cutting, system for continuous irrigation solution
1926	Stern	The first resectoscope for cutting prostatic tissue under visual inspection
1928	Davis, Bovie	Wappler's tube generator was added to generators producing sparks for coagulation in the Davies-Bovie-generator. Principles still used today.
1954	Hopkins	Light transmitters of glass fibres and rod lenses
1970s		Semiconductors in diathermy generators
1980s		Computer chip video camera

Table 2. History of transurethral resection of the prostate

Training the surgeon

The cornerstone of surgical training is the traditional apprenticeship model introduced by William Halsted[1]. This model consists of three key components: observation, coaching and practise and lies behind the saying – "see one, do one, teach one". These training principles have been used basically unchanged ever since. In an article by Wanzel et al[16] many of the modern theories and aspects of acquisition of surgical skills are discussed and one ***motor skill theory*** with many similarities of the apprenticeship model was described by Fitts et al in 1967[17]. They suggested that motor skills are learned in three major stages: cognitive, integrative and autonomous. During the ***cognitive phase***, the learner intellectualises the task into its component steps by reading, listening and watching the new procedure. It is an identification and development of the components which involves the formation of a mental picture and an executive programme of the skill. Performance during this phase is erratic and the procedure is carried out in distinct steps. In the ***associative phase*** the components are linked into a smooth action by practising the skill using structured feedback. The knowledge of the components is integrated into appropriate motor behaviours – the executive programme is practised and performance becomes more fluid with fewer interruptions. During the final stage, the ***autonomous phase***, the skill develops to become automatic. It involves little or no conscious effort. Not all performers reach this stage. The learning of physical skills requires the relevant movements to be assembled, component by component, using feedback to shape and polish them into a smooth action. Rehearsal of the skill must be done regularly and correctly.

Another prominent theory is the ***schema theory of discrete motor skill learning*** developed by Schmidt 1975[18]. Schmidt argued that individuals do not learn specific movements but instead construct "generalised motor programmes." They do this by exploring programming rules, learning the ways in which certain classes of movement are related. They then learn how to produce different movements within a class by varying the parameters that determine the way in which movements are constructed**.** As people practise a movement, such as throwing a ball various distances or in various directions, or climbing stairs of various dimensions, they learn the relationship between movement parameters and outcome. By collecting "data points" they improve their understanding of the relationship between a movement outcome and their control of the movement's parameters. Schmidt's schema is based on the theory that that every time a movement is conducted four pieces of information are gathered in two "schema", hence the name:

Recall schema

- Initial conditions - starting point – information about position and environment from various receptors (e.g. proprioceptive, visual, auditory) which helps to plan the action.
- Response specification - how fast, how high – generation of specific muscle commands

Recognition schema

- Sensory consequences – response-produced sensory information - This information consists of the actual feedback stimuli received from the eyes, ears, proprioceptors, etc. Thus, the sensory consequences are an exact copy of the afferent information provided on the response.
- Response outcome – the success of the response in relation to the outcome originally intended. The desired outcome of the movement is potentially a verbalization, such as, "put the stitch in the centre of this area", and the response outcome is in these same terms, such as, "you put the stitch 4 mm to the left". Thus, the actual outcome of the movement is stored, not what was intended. The accuracy of the outcome information is thus a direct function of the amount and fidelity of the feedback information and a *subject without any feedback information does not have outcome information to store.*

An important prediction of the theory is that the student will more quickly learn the relationship between manipulating parameters and achieving a desired movement outcome if they practise a task in a wide variety of situations and experience errors in the process. Structured feedback closely following performance is also important. In a study by Ahlberg and co-workers on the performance of inexperienced surgeon's in their first laparoscopic fundoplications, a variation in learning curves was shown and the supervisor was the most important factor influencing the inexperienced surgeons performance score[19]. In another study by Kruglikova and co-workers on performance in VR-simulated colonoscopy it was shown that the group receiving structured feedback from an experienced supervisor showed a steeper learning curve with fewer errors than the group training with simulator-generated feedback only[20].

Theories about superior or expert performance, which lie at the other end of the spectrum, may also provide insight into the conditions optimal for surgical training. Ericsson[21] conducted research on the acquisition of expertise in sports, music and other professions. His "***ten-year rule of necessary preparation***" claims that it takes at least a decade of deliberate training to acquire expert knowledge and technical skills. According to this rule, not even the most talented individuals can attain international performance without approximately 10 years of preparation; the majority of international-level performers have spent considerably longer. The actual time of experience with relevant activities is only weakly related to performance. An important reason for this weak relation is that many of our most common activities afford few opportunities for effective learning and skills improvement. Ericsson used the term "***deliberate practise*"** for the individualised training activities designed by a teacher to improve aspects of an individual's performance. To receive maximal benefit from feedback, individuals have to monitor fully concentrated training, which is effortful and limits the duration of daily training. Ericsson also argues that "practise without full concentration may actually impair rather than improve performance". An analysis of these performers' daily patterns of practise and rest indicated that the maximal amount of fully concentrated training that they could sustain every day for years without leading to exhaustion and burn-out was around four hours a day. Ericsson challenges the common belief that exceptional achievements reflect unique abilities or an ***innate*** talent. He argues that the influence of innate talent on expert performance is small or even negligible. Instead the motivational factors that predispose individuals to engage in deliberate practise are more likely to predict differences in levels of expert performance. Complex cognitive skills, such as improved memory, can be acquired through deliberate practise[21,22]. Ericsson also suggests that training with medical simulators may incorporate the characteristics of deliberate practise[23]. In a review of the use of high-fidelity medical simulators it was shown that the hours of practise have a strong empirical association with standardised learning outcomes[24].

There are forces that affect the acquisition of surgical skills negatively. There is the pressure to keep a high turnover in the OR and there are substantial costs associated with learning in the OR. A calculation of the cost of total operative time lost when residents performed the procedure was made by Bridges et al in 1999[25]. During a residents four years of training the extra cost was almost $50 000. The complexity of surgical practise has increased with ever increasing medical knowledge and introduction of new techniques which has led to the development of surgical sub-specialties and "super-specialists". The teaching hospitals are more and more populated with patients with serious and complex surgical problems that demand the skill of experts, the less complex procedures are located to high-turnover, specialised units. Learning disease and operative technique by random chance and opportunity is becoming increasingly difficult[26]. Finally, public expectations and ethical considerations make it unacceptable to learn basic techniques on real patients.

The OR as the venue for the practise of surgical skills is questionable. There is a need to develop possibilities to learn surgical skills outside the OR. As early as 1962 a course in surgical technique was held in Canada[27] and was followed by others during the 1970s [28]. A review of the surgical education literature published between January 1988 and august 1998 reflect the change in attitude to teaching technical skills outside the OR. During the first two years of that period there were no studies, but this was followed by a growing interest during the latter half of the decade[29]. Today there are many studies published in this area and the development is partly due to the advancement of reliable and validated assessment tools to measure surgical skills.

Measuring surgical behaviour

It is difficult to define what makes a skilful surgeon[30]. Surgical competence is a mixture of technical skills, good judgment, commitment and patient concern. All these ingredients can be further subdivided but it is still difficult to recognise which mix is the most beneficial. Good judgment must be based on knowledge which can be assessed through theoretical exams. To assess how well the resident uses this knowledge in the clinical situation e.g. case-based scenarios and videotaped patient encounters can be analysed. A common way of evaluating a resident's progress is ***In-Training Evaluation Reports*** (ITERs) which is an ongoing assessment of the resident during day-to-day work. It is usually composed of global rating scales assessing multiple dimensions of competence, including technical skills. There are indications that ITER is poor at identifying residents with poor technical skills[31]. Technical skills are important in surgery and despite this obvious fact these skills have been poorly evaluated during surgical training programmes. It is common to use a logbook listing clinical and surgical experiences - type of operation and if the resident was primary surgeon or not. This logbook is in many countries a requirement for licensure. A more modern approach to the logbook is ***Cumulative Sum Analysis*** (CUSUM). This is a statistical tool based on a logbook where variables such as success/failure rate are recorded. Acceptable failure rate in for instance cystoscopy can be set at 10 per cent. Each successful procedure decreases the value with 0.1 and a failure adds 0.9. The trainee is competent when the trend falls and remains below a boundary of 0.9. This will identify residents with persisting difficulties and also whether or not the training programme provides enough exposure to a particular procedure. Both these methods give information about quantity but rather poor information about quality. There is a growing concern about assessing technical skills in a more objective manner and there is increasing evidence supporting well-validated objective assessment methods[32]. The best available way of assessing technical skills appears to be ***observation with criteria***. Observation can be direct, with the assessor physically present or indirect using video-taped performances. The most used direct method to evaluate surgical technical skills is the University of Toronto's ***OSATS***-model (***Objective Structured Analysis of Technical Skills***)[33-39]. The assessment instruments are similar to those developed for OSCE (Objective Structured Clinical Examination), developed by Harden et al[40] in 1975, which is

• Live operation – the "gold standard" of assessment but raises many practical and ethical considerations. • Human cadavers – ethical considerations and also problem with the variability of tissue. • Animal models – anatomical and ethical disadvantages. • Bench models / Box trainers – possibility to train a great variety of skills but gives no metrics or feedback • Computer-based trainers – Same advantages as box/bench trainer but with metrics and automated feedback.

Table 3 Different models used when assessing technical skills.

well-known and widely used. The evaluation criteria come from a checklist and a global rating scale. The checklist is detailed, procedure-specific where one point is given if the item is performed and no point if performed incorrectly or not at all (check of patient id before operation starts – yes/no). Global rating scales consists of multiple items, each rated on a behavioural scale (Flow of operation – 1: frequent stops, 3: reasonable progress, 5: effortless flow). OSATS has a rather good reliability [33,34,41] and also proof of validity[41]. The method is objective and can be applied in all environments and models (Table 3) but requires an extensive amount of the expert examiners time and hence high costs. The indirect method of using previously video-taped performances[19,42-44] puts fewer demands on logistics and costs and can even be done in "fast-forward" with preserved reliability.[45]

An assessment instrument must to be reliable and valid. ***Reliability*** is the consistency or repeatability of the measures. There are different classes of reliability estimates:

- Inter-rater reliability – different assessors, same assessment instrument.
- Inter-method reliability – different assessment instruments, same "target".

- Test-retest reliability – same assessor, same assessment instrument, different assessments. Assessments made on two separate days of the same individual shall give the same results.
- Internal consistency reliability – consistency of results across items within a test. Used to check that the test person gives the same answer to questions regarding the same subject.

However, reliability does not imply validity.

Validity

Validity - the extent to which an assessment measures what it is meant to measure and the extent to which inferences and actions made on the basis of test scores are appropriate and accurate. It is also a measure of usefulness, meaningfulness and appropriateness. Validity is not a simple notion, it is compromised of a number of principles and a number of validation benchmarks have been developed to assess the validity of a testing instrument. It is important to take the consequences of the use of a test into account. *"Social consequences of testing may be either positive , such as improved educational policies based on international comparisons of student performance, or negative, especially when associated with bias in scoring and interpretation or unfairness in test use"*.[46] The benchmarks used in this thesis were defined in a well-sited article by Gallagher et al from 2003, including the concepts of face, content, construct, concurrent, discriminate and predictive validity[47]. The same concept was recommended in the European Association of Endoscopic Surgeons (EAES) consensus guidelines[48] 2005:

Face validity evaluates whether or not the test is appropriate and if it "looks like" it will measure what it is supposed to measure. This is a subjective form of validation and is performed by experts during the initial phase of test construction.

Content validity is the systematic examination of the test content to determine whether it covers a representative sample of the domain to be measured. It is based on a detailed examination of the test items content. Does the test contain the steps and skills that are used in the procedure? This is also a subjective process which relies on the opinion and judgment of experts.

Construct validity seeks agreement between the theoretical concept and the specific measuring device or procedure. Can the test items identify the quality, ability or trait it was designed to measure? One example is the ability to differentiate between experts and novices performing a given task.

Concurrent validity is demonstrated where a test correlates well with a measure that has previously been validated. This is a measure of agreement between the results obtained by the given survey instrument and the results obtained for the same population by another instrument acknowledged as the "gold standard".

Discriminant validity is the degree to which a test score does not correlate with scores from other tests that are not designed to assess the same construct. In validity tests of medical simulators the term *discriminate validity* is used meaning "an evaluation that reflects the extent to which the scores generated by the assessment tool actually correlate with the factors with which they should correlate"[47]. It is in these cases used to differentiate ability levels within a group with similar experiences.

Predictive validity is the agreement between results obtained by the evaluated instrument and results obtained from more direct and objective measurements.

Simulators in surgery

Simulation is the replacement of a potentially dangerous procedure in the operating room by the enactment of a similar procedure in a simulated environment. Simulation in surgery has a long history in the form of human cadavers and animal models[49]. The drawbacks with the use of simulation include cost, limited availability, non-compliance of tissue, specialised facilities and ethical concerns. For instance, the use of animals for teaching surgical skills has been banned in the UK since 1876[16]. Inanimate models have several advantages; they are safe, portable, reproducible and cost-effective. Low-fidelity bench models may be as useful as the more sophisticated high-fidelity VR-simulators[50-53], but lack the possibility of objective performance measures and automated feedback. Teaching, rehearsal and assessment occur simultaneously when using a computer-based simulator. The concept of VR-simulation in surgery has only a 20-year history. The first commercial surgical simulator was MIST-VR, which combined a mechanical "box-trainer" with an abstract graphic image. The first documentation by Sutton et al. in 1997[54] describes a simulator where fundamental skills and tasks could be trained, such as pick and place, transfer of objects. The value of this simulator was demonstrated by Seymour et al.[43] by documenting a reduction in operation time by 29% and a decrease in errors by 85% on gallbladder dissection in cholecystectomy. There are nowadays a vast number of simulators designed for different tasks[55]. The more advanced simulators include the use of the actual instruments, a virtual image of the appropriate anatomy is displayed and interaction with virtual anatomy and pathology is possible. One of the most sophisticated is the Vascular Intervention Simulation Trainer (VIST) which is used for endovascular procedures[56]. This simulator can import patient-specific CT scans and the surgeon can practise on the actual patient data before performing the procedure. The capabilities of this simulator have led the Food and Drug Administration (FDA) in USA to suggest that virtual reality simulation would be an important component of a training package for carotid stenting[57].

When it comes to specific urological simulators, there are several commercial applications of which most have gone through validation studies. There are modules applied to laparoscopic simulators for rehearsal of specific urologic procedures such as nefrectomy and also specific endourological simulators such as for training ureterorenoscopy. In a qualitative and systematic review of training models in endourology by Schout et al in 2008 [58] 45 articles were found, describing 30 models for endourological procedures. Of these, only three were classified as randomised controlled trials (RCT). The most common validation study was for ureterorenoscopy (26) and the least common (1) was for transurethral resection of bladder tumours (TURB). Eleven studies contained five models for TURP and seven of these were validation studies. Another finding was that it was university departments that developed the TURP models without involvement of commercial companies. Commercial companies' interests may not always coincide with urologists' educational goals. It was not possible to do any statistical analysis of these studies because of the low number, low level of evidence and too few RCTs (Table 4).

The important question in training surgery outside the OR is whether the skills gained in a simulated environment translate to improved performance in the OR – "VR to OR". A systematic review of virtual reality simulators for training laparoscopic surgery was done by Gurusamy et al in 2009 [59] who investigated the effect of skill acquisition by training in VR-simulators. They found 23 trials with 622 participants that fulfilled the criteria to be included in the analysis. The conclusion was that VR training can supplement standard laparoscopic training and that it is at least as effective as video training (i.e. box trainer). In a review by Seymour in 2008 [60], 14 studies of VR to OR was included in which 10 studies analysed the effect on operations on humans and four on pigs. Of the seven studies on laparoscopic skill transfer, one failed to demonstrate skill transfer. Of these seven, only three used surgical residents as study objects. He argues that the ethical question is a major concern when using "non-training" control subjects when evaluating procedures on humans. It will also be problematic to extrapolate training results from medical students to training residents. Skill transfer from VR for surgical residents requires assessment of both residents and expert clinical performance on tasks that may be inappropriate for a medical student. In the EAES consensus guidelines on validation of VR-simulators[48] it is defined that a randomized trial is the highest qualitative level of evidence. However, as Seymour states, it may be better to evaluate if the training curriculum using simulators is effective or not. There is a risk that the manufacturers of simulators develop their simulators to contain the necessary steps and procedures that are suitable for validity tests including inexperienced study subjects instead of the procedures necessary to improve the performance of surgical residents.

Author	Publ. year	Manufacturer	Validation	Remarks
Trindade et al[49]	1981	Univ. California, US	-	Animal model
Ballaro et al.[61]	1999	University College, London, UK	Content	Low resolution model, no haptics
Kumar et al[62]	2002	Imperial College of Sci, Technol. and Med., London, UK	-	Physical model with superimposed VR-information
Sweet et al[63]	2002	Univ. Washington, Seattle, US	-	Rudimentary design
Sweet et al[64]	2004	Univ. Washington, Seattle, US	Face, content, construct	Large number of participants, 5 minute assessment time
Källström et al[65]	2005	Melerit AB, Linköping, Sweden	Face, Content	Preliminary construct validation
Rashid et al[66]	2007	METI Surg. Sim. TURP	Discriminate	Same study group and simulator as in Sweet et al[64]
Padilla et al[67]	2007	Univ. Nacional Autonoma de Mexico	-	Rudimentary design
Bach et al[68]	2009	Asklepios Hospital Hamburg, Germany.	-	Low-fidelity, home-made "box-trainer"
Schout et al[69]	2009	Karl Storz GmbH, Tuttlingen, Germany	Face, content	Modification was recommended before initiating further experimental validity studies.
Källström et al[70]	2010	Melerit AB, Linköping, Sweden	Construct	Small number of participants, 3 and 6 full procedures assessed

Table 4. Studies of simulators for the training of transurethral resection of the prostate

Learning curve

There is a gap in empiric research between studies that focus on improved performance on simple tasks during a single, short session and studies of expert performance after years of deliberate practise. Studies of learning curves may address this gap. The learning curve is the relationship between experience with a procedure and outcome variables, such as procedural time or complication rate. Improvement occurs more rapidly during early experience which makes the early part of the curve steep. Gradually, less and less new information is retained after each repetition and the curve evens out.[71] With the "correct" outcome variables it would be possible to define when the individual reaches proficiency. It would also be helpful in a surgical curriculum to define the limit for when additional training gives very little improvement. Analysis of performance curves may lead to better understanding of the variables that influence the learning and provide guidance for the development of surgical curricula[72].

Errors

The original Hippocratic Oath (from the 4th century BC) taken by doctors worldwide, states "I will use those dietary regimens which will benefit my patients according to my greatest ability and judgement, and I will do no harm or injustice to them", but later in the oath it is also stated "I will not use the knife, even upon those suffering from stones, but I will leave this to those who are trained in this craft"[73]. This could be interpreted in two ways; the first is that it is impossible for a single person to maintain expertise in all areas, but it could also be interpreted as that it is not possible to perform surgery without doing harm. From the "To err is human" report in 1999 it is estimated that 44 000-98 000 patients die annually in the USA due to medical errors[74]. This makes it the eighth most common cause of death in USA. In Sweden the number of adverse events in healthcare is estimated to between 3-16%. Investigations into the nature of error have given a new understanding of the cause and effect. Most errors are "systemic" in nature; the errors occur within an entire system of events, a single noticeable event is often not the only cause of the error. In 2005 WHO established the World Alliance for Patient Safety that states *"Current conceptual thinking on the safety of patients places the prime responsibility for adverse events on deficiencies in system design, organisation and operation rather than on individual providers or individual products*"[75]. Although this is a very important element it may remove too much blame from the individual and lead into avoiding responsibility[76]. In surgery there is often a specific action that, in itself, is the error. This is the active, specific error where the surgeon is responsible, but there is also an opposing point, the general, latent error where "the system" has responsibility. A single error lies somewhere on the line between the two. There are many possible systemic causes of surgical errors in i.e. the diagnostic workup (delays, outmoded or failing tests), delayed treatment (lack of resources), outmoded or failing equipment. But there are also errors that depend on the individual (faulty conclusions, choosing sub-optimal or wrong treatment, error in performance of an procedure, poor communication skills). These errors are partly of a systemic nature (poor instructions, education or knowledge of the individual surgeon's abilities) but it is also the individual's responsibility to understand his/her limitations if errors are to be avoided.

Psychometrics

Neuropsychological factors are important when it comes to learning and studies focusing on surgical skills have found that visuospatial ability and visual working memory correlates positively with performance measures in surgical procedures[77-80]. Manual dexterity seems to be a poor predictor of surgical skill[81-84] and in studies of manual dexterity results contradict the surgical folklore that pure motor skill predicts surgical performance, instead it seems that visuospatial ability is more important[78,85-87]. In image-guided surgery, such as TURP, visuospatial ability is believed to be even more important than in open surgical techniques[88].

Visuospatial ability refers to the individual's ability to generate a mental representation of a two- or three-dimensional structure, assessing its properties and performing a transformation of its representation[89]. Visuospatial ability comprises multiple distinct, but interrelated subcomponents. There is a gender difference in visuospatial ability suggesting that women use strategies different to men[90]. Men have on average higher scores on visuospatial tests and in women spatial ability correlates with verbal ability whereas not in men[91]. There is also a tendency for older individuals to perform worse than younger adults, even amongst people who frequently use visuospatial abilities in their profession[92]. The ability to see an object as a set of parts and then be able to construct a model of the original from these different parts is known as the ***constructive aspect of visuospatial ability***. This is a central cognitive ability that includes combining parts into a meaningful whole, distinguishing right from left, discriminating between objects, understanding how objects relate to each other in space, adopting various perspectives and to represent and rotate objects mentally[93]. Visuospatial constructional ability also includes understanding and interpreting symbolic representations of external space and the ability to work out the solution for non-verbal problems. ***Working memory*** is often regarded as fundamental function underlying other executive functions. In surgery, and especially during image-guided surgery, with delays, multitasking and interruptions, working memory is important. There are many distractions during surgery and in a study by Healey et al there were on average 0.3 distractions per minute[94]. The ***visual working memory*** is part of a three-part system for retaining *and* manipulating temporary task-relevant information[95]**.** The system contains a ***central executive mechanism*** controlling both the visual working memory, that involves storing and retrieving previously experienced visual or visuospatial information when the original stimuli are no longer present, and the ***verbal working***

memory, that acoustically codes information. Although previous research has found a correlation between surgical performance and visual but not verbal working memory[96], verbal working memory may be of importance when acoustic or verbal feedback is a part of the surgical learning situation. These neuropsychological factors alone, however, do not account for surgical skills. Cognitive processes have to be integrated with knowledge, personality traits and experience and then regulated in order to achieve a future goal. This complex process is often referred to as ***executive functioning***. Though not a unitary construct, executive functioning/planning can be broadly defined as a cognitive domain that involves the delineation, organisation, and integration of behaviours needed to achieve a goal[97]. Visuospatial abilities, working memory and executive functioning can be readily assessed using empirically validated and standardised psychometric instruments[80].

Personality

Several reports address the question of a specific surgical personality as a group [98-100], but do not describe specifically the personality of the good and skilled surgeon. It is relatively easy to describe the "right stuff" military pilot, but despite several studies on military aviators no consensus is made among psychologists on the personality associated with success in aviation[100]. The attrition rate from undergraduate pilot training is about 20% (about the same as for surgical residents) and is believed to be caused of poor motivation and not aptitude[101]. This is in line with Ericsson's theory that ***deliberate practise,*** during the decade it takes to become a skilled professional, mostly depends on motivation. It is not necessarily the general personality traits associated with surgeons as a group which predicts success. A negative correlation, for example, has been shown between self-belief and surgical skills[102]. It is also common for a profession to preserve the standing of their profession by selecting trainees similar to the peers[103]. The trend towards a higher degree of specialisation makes it important to find the "suitable personalities" but even more to guide the aspirants towards another career if the ability to become a skilled specialised surgeon seems low.

Personality is a complex concept consisting of thoughts, emotions and behaviour. There are many theories today about personality but no consensus on its definition. One of the major approaches is the ***trait theory***. Traits can be defined as habitual patterns of behaviour, thought and emotions which are relatively stable over time, differ between individuals and influence behaviour. Gordon Allport was one of the pioneers in the trait theory. He made a distinction among ***central traits*** (basic of the personality) and more peripheral ***secondary traits*** and where ***cardinal traits*** are strongly recognised in the individual. ***Common traits*** are recognised within a culture and vary between cultures. The ***five-factor model***[104] was developed by the statistical technique of factor analysis performed on objective measures of all known personality traits (from dictionaries etc). The resulting factors are openness (sometimes called intellect), conscientiousness, extraversion, agreeableness and neuroticism (sometimes called emotional stability). When examining the relationship to personality disorders in the DSM-IV (diagnostic and statistical manual of mental disorders) a unique and predictable five-factor profile was found for each disorder[105]. Cloninger, in the mid 1980s, developed a similar and general model of temperament based on genetic, neurobiological and neuropharmacological data. Until now Cloninger's Temperament and Character Inventory (TCI) has been used in hundreds of peer-reviewed publications with reproducible findings in areas including genetics, neurobiology, learning and clinical psychopathology[106]. Furthermore, the validity of the TCI dimensions have been evaluated by comparison with other models of personality including various forms of the five-factor model[107]. According to Cloninger ***temperament*** consists of four different dimensions each of which is 50% to 65% inheritable and appears stable throughout life regardless of culture or social learning [108,109]. Three temperament traits related to the immediate responses of human beings to basic stimuli were proposed: harm avoidance (HA), novelty seeking (NS) and reward dependence (RD). Reward dependence initially included persistence (PS) as a facet. However several studies showed that persistence is actually an independently heritable trait. As a result, persistence is now considered as the fourth temperament dimension. Temperament refers to individual differences in the sensitivity to specific environmental stimuli and the behavioural responses to those stimuli[110]. These responses are, however, under the control of the person's ***character***. There are three different character scales and together they refer to individual differences in self-object relationships, which develop in a stage-like manner through interactions among temperament, family environment, and life experiences[111] (Table 5).

Cloninger suggests that the temperaments novelty seeking, harm avoidance and reward dependence are correlated with dopaminergic, serotonergic and noradrenergic activity, respectively[112]. The TCI is a widely used, reproducible and valid tool used to measure the seven dimensions of personality [109]. The inventory is based on "cross-fostering" analysis of children separated from their parents at birth that provided strong evidence for the contribution of both genetic and environmental influences to behaviour and disorders. Overall, the importance of character exceeds that of temperament in the learning situation, although the importance of temperament factors increase when character function is reduced. This is the case when people are exposed to hunger, tiredness, complications during surgery and other stressful events and it is shown that "stress tolerance" is predictive of operative skill[78].

To create an optimal situation for acquisition of surgical skills a simulated environment which excludes adverse events for the patient is necessary. According to modern theories of skill acquisition there are three phases in the learning process[17]. The first phase is the cognitive understanding of the task – modelling (imagery and mental practise)[113]. The second phase is the deliberate practise which requires

Personality dimension	High Scorers	Low scorers	No. of items
Temperament – emotional responses to stimuli			
Novelty seeking NS	Impulsivity (incentive to approach or initiate behaviour)		40
Exploratory excitability NS1 Impulsiveness NS2 Extravagance NS3 Disorderliness NS4	Exploratory, curious Impulsive Extravagant, enthusiastic Disorderly	Indifferent, rigid Reflective Reserved, detached Orderly, regimented	11 10 9 10
Harm avoidance HA	Anxiety proneness (inhibition of behaviour)		35
Anticipatory worry HA1 Fear of uncertainty HA2 Shyness HA3 Fatigability HA4	Worrying, pessimistic Fearful, doubtful Shy Fatigable, asthenia	Relaxed, optimistic Bold, confident Gregarious Vigorous	11 7 8 9
Reward dependence RD	Sociability (sensitivity to signals of social approval)		24
Sentimentality RD1 Attachment RD3 Dependence RD4	Sentimental, warm Dedicated, attached Dependent	Practical, insensitive Withdrawn, detached Independent	10 8 6
Persistence PS	Perseverance (resistance to extinction of behaviour)		8
	Hard working, ambitious, overachiever, perfectionist	Irresoluteness, modest, underachiever, pragmatic	
Character – the "self-concept"			
Self-Directedness SD	Awareness of being an autonomous individual, "willpower"		44
Responsibility SD1 Purposeful SD2 Resourcefulness SD3 Self-acceptance SD4 Congruent second nature SD5	Responsible, reliable Purposeful Resourceful, effective Self-accepting Mature, strong	Blaming, unreliable Lack of goal direction Inert, apathy Self-striving Immature, personal distrust	8 8 5 11 12
Cooperativeness CO	Recognising self as an integral part of society		42
Social acceptance CO1 Empathy CO2 Helpfulness CO3 Compassion CO4 Pure-hearted conscience CO5	Socially tolerant Empathic Helpful Compassionate, constructive Ethical, principled	Intolerant Social disinterest, Critical Unhelpful Revengeful, destructive Opportunistic, self-serving	8 7 8 10 9
Self-Transcendence ST	Participation in the world as a whole		33
Self-forgetful ST1 Transpersonal identification ST2 Spiritual acceptance ST3	Creative, self forgetful Wise, patient United with universe	Self-conscious Self-isolation, Impatient Rational materialism	11 9 13

Table 5. The scales and subscales of TCI-R, the revised Swedish version of the Temperament and Character Inventory[114].

structured feedback in close correlation to the performance (no feedback – no learning)[18,20,21,115,116] and where repeated execution of the skill is the key.[18,21] This is also true for the third phase, the autonomous phase, where skills become automatic and where rehearsal must be done regularly and correctly. It is important to practise the task in a variety of situations and experience errors in the process. The clinical situation in teaching hospitals cannot provide this due to practical and ethical reasons. Moderate tension levels often enhance learning due to an increased level of arousal, but this effect is reversed if the level of tension creates anxiety.[117,118] The operating room environment is often too stressful to create good learning conditions. A well-designed and validated computer-based VR simulator, including automated structured feedback and metrics, can satisfy most of the criteria required for a good learning environment but does not replace the mentor or the curriculum[52]. Used as a learning tool amongst others, it may help to reduce adverse events for patients and shorten the learning curve in the OR.

It is estimated that about 5-10% of trainees do not possess the innate abilities that are necessary to reach proficiency in image-guided surgery and these skills may not improve with practise[119-121]. This may not be true; according to Ericsson the important quality to acquire skills is motivation. It is also shown that abilities proposed to be innate, such as visuospatial ability, can be improved by systematic training and/or computer gaming experience[83,122-124]. Nevertheless, it is of great value to reveal underperformance early, before the decision of specialisation. Early evaluation of abilities (or, according to Ericsson: motivation) would allow for further training or career guidance towards other less practical specialties. Furthermore the surgical profession needs a reliable and valid method of assessing the skill of its members[125,126], especially when new techniques are introduced. A "driving test" may not be a guarantee against errors but it makes them less likely to occur.[127]

A simulated procedure is a limited environment, without many of the stress-factors present in a real operation room. Stress vulnerability differs between individuals as well as the coping mechanisms to manage the reactions it raises. It is therefore important to estimate these personality traits to be able to foresee a person's behaviour in the OR. Temperament is mostly inherited and can only be affected by experience to a certain degree. Deficiencies in the temperament can be compensated by traits in the person's character – which can be affected by learning and experiences. The balance in temperament and character provides the conditions for performing well. If the character needs to compensate for many or severe deficiencies in the temperament the compensatory effect of the character may not be enough. When the person becomes tired, scared, stressed, hungry etc, the ability to compensate falls and the deficiencies become visible. It is of value for all to be aware of this balance and to be careful not to exceed the personal limits. It may also be helpful for the individual to know about his/her deficiencies in character traits since they can be improved by experience.

Aims

The general aim of this thesis was to construct and investigate the validity and use of a virtual reality, real-time, simulator for the training of TURP. This was divided into the following aims:

- To discover the requirements for and to construct a simulator for the training of transurethral resection of the prostate (Face and Content validity)
- To evaluate the learning curves for inexperienced and experienced performers and the differences between the groups (Construct validity)
- To design an effective training program including a TURP-simulator
- To evaluate if practising the TURP procedure in a VR-simulator increases the skills and dexterity of urology residents when performing the procedure on patients (VR to OR) without increasing the risk for patients
- To find out if implementation of simulation technology can be recommended in the general urological education curriculum
- To evaluate if there is a specific urological/surgical personality.
- To evaluate if it is possible to predict future surgical skills using evaluations of performance in a simulated environment together with results from personality and psychometric tests.

Materials and methods

Face validity

A rudimentary demonstration version of the simulator was presented to a group of experienced urologists and afterwards they were asked to answer a questionnaire regarding training experiences (Table 6).

List the three most common errors an inexperienced urologist makes during his/her first TURPs?
What are the most difficult problems according to the inexperienced urologist?
What was most difficult for you when you started to perform TURP?
Which part of the procedure do you still find difficult?
Do you think that inexperienced urologists may benefit from quantity training on a TURP simulator?
What parts of the procedure must be represented in a simulator programme?
What aspects of the graphical presentation must be present in a simulator programme?
Do you think that you may benefit from quantity training or practise difficult situations on a simulator?

Table 6. Face validity questionnaire

Design of the simulator

In cooperation with Melerit AB (Linköping, Sweden) a full-procedure simulator for training in TURP (Pelvic Vision) was constructed. The development was based on an iterative process with testing, improvements and re-testing until a "final" version was accomplished. This version was used for the trials in this thesis. Between Papers I and II forces were added in the Z-direction when moving the resectoscope and there was an upgrading of the graphical interface mainly affecting the user-interface and the number of polygons used in the graphical and haptic model. The assessment measurements were not changed.

Hardware

The hardware consists of a desktop computer (two Intel Pentium 3 processors, 800 MHz, 256 MB RAM, Windows 2000, ASUS Geforce III graphic card) and a monitor (1280x1024 bpi). A robotic arm system (SenseAble Phantom Premium 1.5) with 6 degrees of freedom for motion control and 3 degrees of freedom for haptic feedback is connected to the system. The robotic arm was connected to a modified resectoscope (sensors in the cutting grip, the stopcock and the connector) controlling the simulation of the cutting loop movement, irrigation fluid flow and disconnection of the instrument. The simulated coagulation/cutting current is controlled via foot pedals (Fig 1), sound alerts are heard and “bubbles” emerge from the cutting loop when pressing the pedal. The resectoscope runs through a fixed point to represent the pivot-point in the pelvic floor. This is created with a table-mounted metal frame with an inner lining of thick rubber to simulate the resistance of the musculature and fasciae of the pelvic floor. There is also a pair of artificial legs to create a more realistic environment (Figure 4).

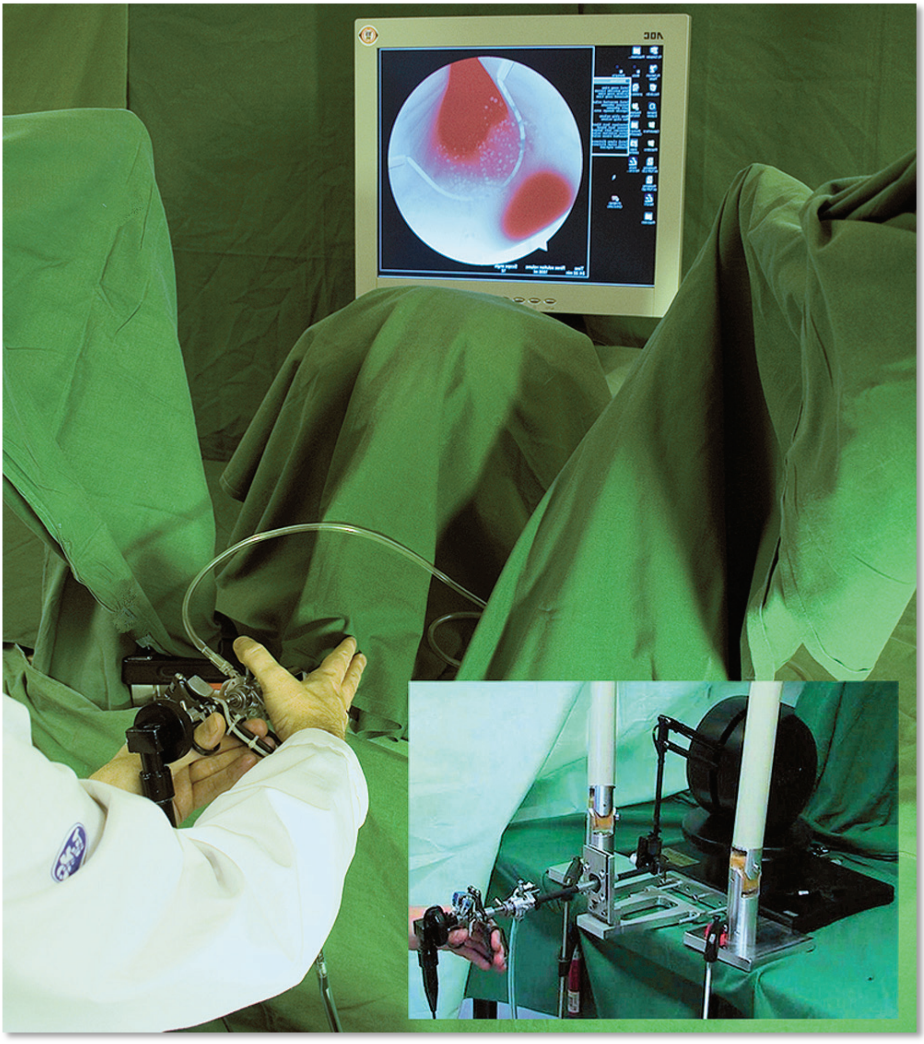

Figure 4. The simulator (PelvicVision) with the modified resectoscope, rubber-lined pivot-point in metal-frame, Senseable® robotic arm system, monitor and “legs”.

Software

Product Name	Company	Description
Melerit M-Base	Melerit Medical AB, Sweden	Software tool to develop medical simulators including deformation calculations
OpenGL	Silicon Graphics	Standardised graphic library
Cosmo3D/Optimizer	Silicon Graphics	High level application programming interface for graphics
GHOST	SensAble Technology	Controlling the robot arm
JAVA SDK	Sun Microsystem	Graphical user interface

Table 7. The software in the PelvicVision simulator.

The software (Table 7) has three working areas:

1) The simulation module which computes the information from the robotic arm, the resectoscope handle, the resectoscope fluid tap and the foot pedal calculates the data fed into 2) and 3);
2) Graphic rendering which generates the virtual view on the screen;
3) 3-D haptic rendering that generates the haptic control of the robot arm.

Figure 5. The view of the prostatic urethra when performing simulated resection. An arterial bleeding is seen on the left side.

The simplified anatomical model has a spherical *"prostate capsule"*, the *"prostatic lumen"* is shaped by bilateral bulging "*adenomas*" (Figure 5). The surgical capsule is marked by a slight change in colour when cutting into it and a perforation is clearly marked by colour. The rest of the "urethra" is cylindrical and the "*verumontanum*" is placed at the distal, dorsal part of the "*prostate*". The angle of the "*prostatic lumen*" in relation to the "*bulbous urethra*" can be changed. The elasticity in the tissue can be predefined as well as the bladder volume and the size of the prostate. The surfaces are generated from photographs of the corresponding anatomy. The various optic angles (0, 10, 12, 30, 70 degrees) may be changed during real-time. Arterial and venous bleeding occurs randomly when cutting the tissue and the volume is calculated and presented on the screen. There are four areas where bleeding is more likely to occur according to the normal anatomy, and the size of these areas as well as the probability of bleeding can be predefined. The blood loss and the visual impression of bleeding depend on the vessel diameter, the blood pressures and the intravesical pressure. Arterial and venous bleeding is illustrated with overlaid film clips, and the bleed causes a general "blood fog" which occludes the vision. A high irrigation flow gives better vision, especially close to the resectoscope tip.

The intravesical pressure is calculated from the size of the bladder and the volume of irrigation fluid instilled in the bladder (Figure 3)[6]. When the volume of irrigation fluid in the bladder gives a pressure exceeding that in the veins the corresponding uptake of irrigation fluid is calculated and presented on the screen. The intravesical pressure also affects the calculated volume and number of visible bleed. When the pressure rises in the bladder the pressure gradient becomes less and the bleeding decreases and eventually stops, both visually and in the summation of calculated blood loss. The effect of the irrigation fluid flow on blood fog, pieces of cut tissue and thereby vision is calculated in a similar manner. The virtual bladder has a predefined capacity and when full, the trainee must empty the bladder in the same manner as in the operating theatre, i.e. the inner part of the resectoscope has to be removed from the outer sheet and the screen becomes black. The "emptying" is presented on the screen as a shrinking bar representing the volume remaining in the bladder. When reconnecting the resectoscope, the view of the actual position is presented. The amount of blood fog is continuously calculated, often giving a "red screen" when the bladder has been emptied. There is also an option when starting the system to use continuous flow irrigation and the trainee does not have to go through this procedure.

Total time of the procedure
Inactive time, that is when not cutting, coagulating or emptying the bladder.
Time with high intravesical pressure
Time spent on cutting and coagulating tissue.
Time pressing the cutting or coagulation foot pedal without contact with the tissue.
Time cutting or coagulating tissue during good vision of the cutting area and time with very poor or no vision.
The total volume of resected tissue and the volume of adenoma still remaining.
The mean and max volume of cut-away chips.
Volume of blood loss.
The number of arteries still bleeding when the trainee has finished.
The volume of irrigation fluid used and absorbed by the patient and the number of times the bladder was emptied.
The total distance travelled by the sling and the resectoscope tip.

Table 8. The parameters assessed when performing a simulated transurethral resection of the prostate

The deformation algorithm is based on a spring-damper model. The volume of tissue cut is presented on the screen and depends on how deep into the tissue the cutting loop is depressed. When coagulating, the loop needs to be close to the source of bleeding but not in absolute contact. The effect is not instant, the current must be applied until the bleeding gradually diminishes and finally stops. In the case-construction module different "*patients*" of different difficulty levels were constructed of which no. 1 was used for the criterion-based practise (Appendix 1). To be able to evaluate the trainee's performance on the simulator, a number of parameters were saved in data files (Table 8).

Content validity

The participants filled in a questionnaire (Table 9) after performing one (Paper I) or three to six (Paper II) TURP procedures in the simulator.

How well does the simulator reproduce difficulties in the following areas?
Orientation, anatomy
Instrument handling
Vision, bleeding
Perforation
Strategy
How do you experience the simulator in the following areas?
Realism
Adaptability
Instrumentation
Overall opinion
Usability in training of inexperienced

Table 9. The questions used in the content validity questionnaire.

Construct validity

The participants were experienced urologists and medical students in the early stage of their education, with no previous experience of surgical procedures. The students attended a short lecture on BPH and the procedure of TURP before starting to practise in the simulator. The urologists were informed about the simulation environment and were asked to proceed as they were used to doing in the operating room. The participants also filled in questionnaires regarding previous experience and their experience of each simulated operation (Appendix 2) after performing the procedure. The urologists performed three and the students performed six complete procedures on three different "patient" cases where the first and last procedures were performed on the same "patient". The student's fourth procedure was performed on the same case as the urologist's second procedure. They were told to take the medical history of the "patient" into consideration to be able to evaluate risk factors.

VR to OR

The concurrent (test agreement with the "gold standard") and discriminant (differentiation of ability levels) validity is combined into the VR to OR test. This is a test of training transfer, i.e. the success of simulator training in actual operative performance.

The course

The study (Paper III) was implemented as a five-day course on treatment of benign enlargement of the prostate. The course included theory, diagnostic methods, the instrumentation used in the TURP procedure, and risk factors. The target group for the evaluation was urology residents in Sweden with some, but minor, experience of transurethral procedures. Three supervised TURP procedures were performed by each of the participants on selected patients. Training was done during supervised VR-simulated TURP procedures until the participant reached at least a predefined "expert"-level (***criterion-based training***) (Table 10). Data from the simulator procedures were saved for later analysis (Appendix 3).

	Trial no.	**1**	**2**	**3**	**…**
	No. of procedures				
Parameter					
Total procedure time	< 852 sec (14.2min)				
Resectoscope tip movement	< 7.76 m				
Inactive time	< 64%				
Resectionvolume/time	> 0.94 ml/min				
Chip size (mean)	> 0.098 ml				
Absorption or Absorption/time	< 0.08 dl				
	< 0.54 ml/min				
High pressure time	< 2.2%				
Bleeding	< 250 ml				
Resected volume	> 13.3 ml				

Table 10. The criterion-based parameters used as a training goal for the urology residents in the construct validity and VR-to-OR studies.

Peroperative assessment

The participants' peroperative performances were rated by one supervisor for each of the two parallel groups. The supervisors had previously gone through the validation forms and had practised the rating procedure together. The validation was done according to the OSATS[33], including a checklist (Appendix 4) and a global assessment form (Appendix 5) with a pass/fail question, and the participants themselves completed a self evaluation form after each procedure (Appendix 2). If the participant failed to perform any of the items on the checklist, the supervisor was told to give the correct instructions immediately. The supervisor was instructed to give instructions when necessary and to take over the procedure if there was any risk for the patient or if the procedure was estimated to take more than 60 minutes. Procedure data (blood loss, resection weight, procedure time) were recorded in the patient charts.

Video assessment

The procedures were video-recorded and analysed "blindly" (without knowledge of the course, participant, supervisor, patient identity or the order of the procedure) at a later stage. Each analysis was made using a minute by minute rating according to an assessment form (Appendix 6). The rating was based on what the main action was during that minute and if this action was successful or not. Serious errors such as suspected perforation of the capsule, damage to the bladder wall, uretheral orifices, sphincter area or urethra were noted. The rating was done by two surgically experienced urologists and the inter-rater agreement (agreements/total number of observations) should be over 90% before analysing independently.

Training transfer

The effect of simulation practise was calculated for each participant as the difference in score for two operations with simulation practise in between, compared to the difference in score for two operations without added simulation practise.

Patient outcome

The patient functional outcome was based on our clinic's regular follow-up at 6-12 months postoperatively including IPSS, bother question, maximum urinary flow rate and incontinence score (Linköping incontinence questionnaire)[128]. Data from all patients included concerning readmissions, reoperations and mortality were obtained from charts from all healthcare facilities in the region.

Predictive validity

The study-group and set-up for the evaluation was the same as in the "VR to OR"-study.

Personality was assessed using the revised Swedish version of the Temperament and Character Inventory (TCI–R), a 238-item, self-administered questionnaire, with measurements of seven separate personality scales. The theoretical assumptions are divided into proposed favourable personality traits that directly or indirectly affect learning in a positive or negative way (Table 11).

	Favourable	Unfavourable
Direct influence	Problem solving skills – TCI: SD3	Poorly balanced impulsivity – TCI: NS2/SD2 > 1
	Orderly – TCI: NS4	Poor impulse control – TCI: NS2>5 and ToLdx: total initiation time < 35 and Total correct score <5
	High Persistence – TCI: PS	
	Confidence – TCI: HA2	
	"Good judgment" – ToLdx: total move score < 16 and total initiation time > 99 (t-values)	
	Visual working memory - Rey CFRT: immediate recall + delayed recall	
	Verbal working memory - WAIS-III: DSF,DSB,LNS	
	Executive planning – ToLdx: Total move score	
	Visual spatial ability – Rey CFRT: Immediate recall + Delayed recall + Recognition	
	Curiosity – TCI: NS1	
Indirect influence	Locus-of-control internalised – TCI: SD1	Low self-acceptance – TCI: SD4
	Goal-directedness – TCI: SD5	High anticipatory worry – TCI: HA1
	Balanced attachment style – TCI: 45 < RD3 < 55 (T-value)	
	High Consciousness – TCI: CO5	
	Careful – TCI: passive-dependent temperament profile, mild type and HA1+HA2 >10 and ToLdx: total problem-solving-time >250 and total move score <10	

Table 11. The theoretical assumptions tested in the predictive validity study in Paper IV.

The psychometric measures (Table 12) were done using Rey Complex Figure and Recognition Trial (RCFRT)[129], Tower of London Drexel University (ToLdx)[130] and the Working memory tasks from WAIS-III[131].The RCFRT was used to measure the subject's visual spatial, constructional functions, and visual memory. The ToLdx was used to assess executive planning and problem solving. The participant's verbal working memory was assessed using the Digit Span subtest and the Letter–Number Sequencing subtest from the WAIS-III test battery. All scores from the analyses of operation and simulation performance were standardised. The raw scores from TCI-R, RCFRT, ToLdx and WAIS-III were used unless stated otherwise.

Rey Complex Figure and Recognition Trial (RCFRT) Was administered according to the protocol described in the professional manual[129]. Each participant is required to draw a complex geometric design with multiple details embedded, first copying with the stimulus present, and then drawing from memory after 3 minutes and again 30 min later. Its demands are multiple and include planning/organisation; visual perception and construction; visual memory encoding, storage, and retrieval processes. The following scores were computed according to the scoring standard established in the professional manual: Copy score, immediate recall score, delayed recall score, and recognition score.
Tower of London Drexel University (ToLdx) The administration, scoring, and interpretation of the measure have been carefully standardised. The participants were instructed to replicate 10 different cylinder patterns presented by the examiner in as few moves as possible while adhering to two specific planning/problem-solving rules. Each participant was allowed 120 seconds to complete each ToLdx test item. The following ToLdx scores were computed: Total move, rule violation, time violation, initiation time, execution time, total problem-solving time. Detailed descriptions of these scores are available elsewhere[130].
Working memory tasks from WAIS-III The administration and scoring followed the detailed instructions in the manual[131]. The Digit span subtest requires each participant to recall digit sequences forward (DSF, two trials per item, 2 to 9 digits) and backward (DSB, two trials per item, 2 to 8 digits). The Letter-Number sequencing subtest (LNS) involves ordering numbers and letters presented in an unordered sequence (two trials per item, 2 to 8 letters and digits). For each part, the test begins with a series of two items presented for recall and continues to a maximum of nine items (eight for DSB and LNS). There are two trials at each series length, and the test is discontinued if the participant fails to remember two trials of the same length. One point is awarded for each correct trial. The maximum possible score for DSF is 16, for DSB 14, and for LNS 21.

Table 12. The psychometric tests used in the predictive validity study in Paper IV.

The analysis of performance in the TURP-procedure and simulation practise was based on the "***learning scores***". Learning score from the TURP-procedures were based on the difference in score between the last and the first operation. The learning score from the criterion-based simulation practise were based on the difference between the first approved and the first performed criterion-based procedure (Table 13).

Video Learning score	From the video analysis the standardised Progress score, Resection score and Independent operation time score were added together and Error score were subtracted.
OSATS Learning score	From the OSATS score the standardised Checklist score, Global score and Pass score (i.e. if the resident were judged to be able to perform the procedure independent from a supervisor) were added and Aborted score (i.e. if the supervisor aborted the resident's procedure due to poor performance) was subtracted.
Operation learning score	Video learning score + OSATS learning score
Simulation Learning score	Resection weight per minute divided by the total practise time in minutes

Table 13. The calculations of learning scores used in paper IV.

On the basis of theoretical assumptions and results from previous studies several hypotheses about correlations between neuropsychological factors, personality and performance in TURP, were developed. Proposed favourable personality traits *directly* affecting learning were e.g.: balanced impulsivity and impulse control, problem-solving skills, orderly, high persistence and low ambivalence. We also proposed a number of traits *indirectly* affecting learning (traits not involved directly in the process but still affecting the person's ability to create a good environment for learning): curiosity, self-acceptance, internalised locus-of-control, goal-directedness, balanced attachment style and high conscience. When it comes to neuropsychological factors we propose that visuospatial constructional ability, visual working memory, and executive planning correlate positively with performance and learning measures. In addition, we propose that high scores in the verbal working memory tasks will correlate with high scores in the learning measure. The theoretical assumptions were tested with the *Operation learning score*.

The participants mean scores from TCI-R[114] and the psychometric tests were compared to the population mean values.

Statistical methods

Paper I: We used ANOVA to compare procedures 1, 2 and 6 and all the students with each other. The difference in performance between procedures 1 and 6 was tested using a single-sided t -test.

Paper II : The statistical analysis of the learning curves was performed using a paired t-test, where the first and the last procedure was the fixed factor and the individuals were the random factor. To check for difference in performance between the groups, analysis of variance was used. The group and procedure number (1,4,6 for students and 1,2,3 for urologists) were fixed factors and the individuals were random factors nested within groups. Analysis of binary data was performed using the Fisher exact test and the McNemar test. All tests were calculated with the statistical software SPSS version 16.0.1 (SPSS inc. Chicago, IL, USA).

Paper III: The statistics used for were the paired samples t-test, sign test (for non-parametric data), and the statistical analysis was done using SPSS version 16.0.2 (SPSS inc. Chicago, IL, USA).

Paper IV: The statistics used were stepwise linear regression, Pearson's correlation test and independent sample t-test. The statistical analysis was made using SPSS version 16.0.2 (SPSS inc. Chicago, IL, USA).

Ethical considerations

The Regional Research Ethics Committee at Linköping University approved the study (M163-06).

Collected data comprise video films recorded through the resectoscope (where the patient cannot be identified) (DVDs), data from simulator training (computer registers), participants' background data (paper) and peroperative assessment (paper), were all id-encoded. Collection of the patient data (peroperative documentation and follow-up data) were pursued according to the clinic's normal procedures.

The risks for the participating residents in urology should not differ from the normal educational situation at their home-clinic. Concerning the risks for the patients the study situation should not differ from the normal clinical situation where residents perform procedures with the help of a guiding tutor. The difference from the normal situation is that there was an accumulation of procedures performed by inexperienced residents in the hospital where the course was held during this period. On a larger scale, it would make no difference for the population since these procedures would have been done in their own hospital anyway. On a larger scale the risks for the patients may be less since the procedures during the study were performed with the help of experienced tutors and in a structured and controlled way. The simulator had previously been validated and considered safe to be used in the teaching of inexperienced residents. The use of a simulator when teaching the procedure may also shorten the learning curve for patient procedures.

A possible violation of the participant’s integrity was the assessments of surgical skills, visuospatial ability, working memory, executive ability and personality. All these assessments were done after informed consent and the participants were also informed of their own results.

Results

Face validity

The answers from 17 experienced urologists (of the 28 that received the questionnaire) could be grouped into five categories in the following order (number of times mentioned in answers in parenthesis):

1. Loss of orientation in the changing anatomy (71)
2. Handling of instruments (46)
3. Bleeding and reduced vision (37)
4. Recognising the surgical capsule of the prostate and perforation of the capsule (28)
5. The overall strategy and operating time (27)

Fifteen specialists believed the trainee could benefit from quantity training in a simulator. Four specialists believed they could benefit from practising with the simulator, six were doubtful and four did not believe so. No one mentioned that tactile feedback was important for a simulator but it was obvious from ongoing discussions that this had been taken for granted.

Based on the results of the questionnaire, emphasis on the development of the simulator was placed on the visual part. Most of the categories could be included in the problem of loss of orientation, especially bleeding that obscures vision. In order to create a realistic environment and evaluate the trainee's performance with the simulator we constructed algorithms for calculating the flow of irrigation fluid as well as arterial and venous bleeding.

Content validity

In Paper I, eleven experienced urologist, in Paper II nine experienced urologists and eleven medical students filled in the questionnaire.

The free text answers from the experienced urologists in Paper I mainly concerned the quality of graphical resolution, and the fact that there were no forces in the Z-axis. The evaluation of content validity in Paper II showed that the students were more satisfied with the simulator's performance regarding vision and instrument compliance. The urologists were more pleased about how the simulator provided the opportunity to practise a strategy. The lowest score was for the simulator's ability to present perforations through the "prostatic capsule". The highest score was for usefulness in training residents in urology. (Figure 6). The acceptability threshold for content validation is set at different levels in different reports[64,69,132]. A score above 6.0 in this set-up could be considered as satisfactory[69].

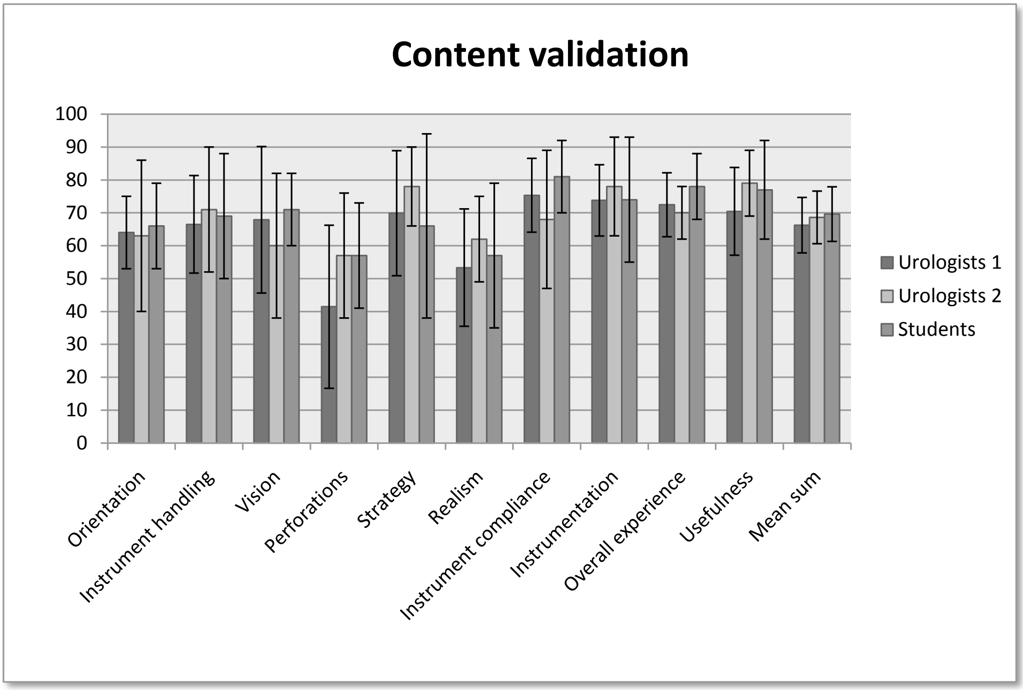

Figure 6. Results from the content validations. Urologists 1 was done by 11 urologists before and Urologists 2 and Students were done by 9 urologists and 11 students after a revision of the simulators contents.

Construct validity

In Paper I a preliminary construct validity test was performed using seven medical students, more data were added and a more thorough validation test was presented in Paper II.

The participants were nine experienced urologists (median 22 years of clinical experience and 900 TURP procedures) and eleven recruited medical students in the early stage of their education (5^{th} term), with no previous experience of surgical procedures.

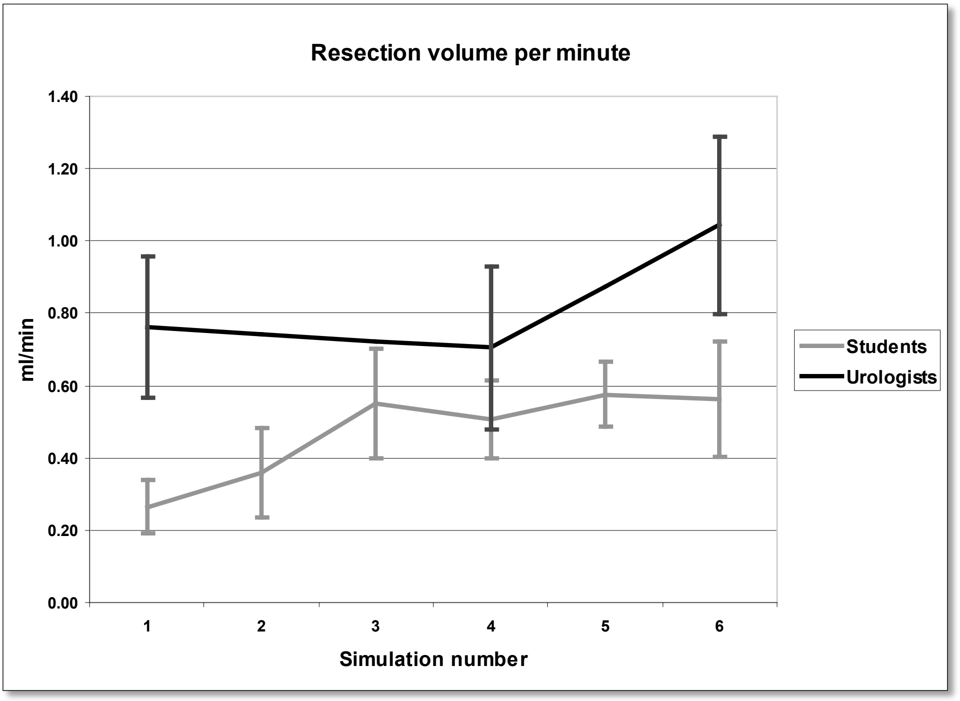

Figure 7.One example on a learning curve measured as resection volume per minute.

The urologists showed significant better results than the students in all parameters except for the amount of blood loss, resection volume and chip size where no differences were found. The urologists caused a blood loss rate that was almost three times as high as the students, but with no difference in total blood loss. There was a positive learning curve for the students in 86% of the parameters but only in 38% for the urologists (Figure 7, Table 14). The participant's inactive time decreased and there was a trend for the amount of time spent resecting to increase. The measured volume of resected adenoma per minute seems to correlate with clinical data, since the experienced urologists reached about 1 g/min in their last operation, which can be compared with 0.8 g/min, reported in other studies.[11,133]

There were significant differences between the group's subjective experiences where the urologists experienced the procedure as easier than the students. The students experience was that the procedure became easier and easier. (Figure 8).

Assessment parameters	Better value	Group Diff	Student learning	Urologists learning
Clinical data				
Total time	U	***	***	**
Resected volume	-	-	***	-
Resected volume/time	U	***	***	***
Completeness	S	-	***	*
Blood loss	S	-	***	-
Blood loss/time	S	***	-	-
Irrigation fluid used	U	*	**	**
Irrigation fluid/time	U	**	-	-
Irrigation fluid uptake	U	**	**	-
Checklist score	U	***	**	*
Serious perforations	U	***	-	-
Simulation data				
Resectoscope movement	U	***	***	*
Resectoscope move/time	U	***	**	-
Inactive time	U	***	***	***
Chip size	U	-	-	-
Adenoma in resected volume	S	***	**	-
High pressure time	U	***	**	*
High pressure time (%)	U	**	*	-
Absorption/resected volume	U	***	***	-
Self assessment data				
Mean self assessment	U	***	***	-
Instrument handling	U	***	*	-
Final result	U	***	***	-

Table 14. Data on construct validity. U - Urologists. S - Students. Significance: * means $p < 0.05$, ** means $p < 0.01$ and * means $p < 0.001$.**

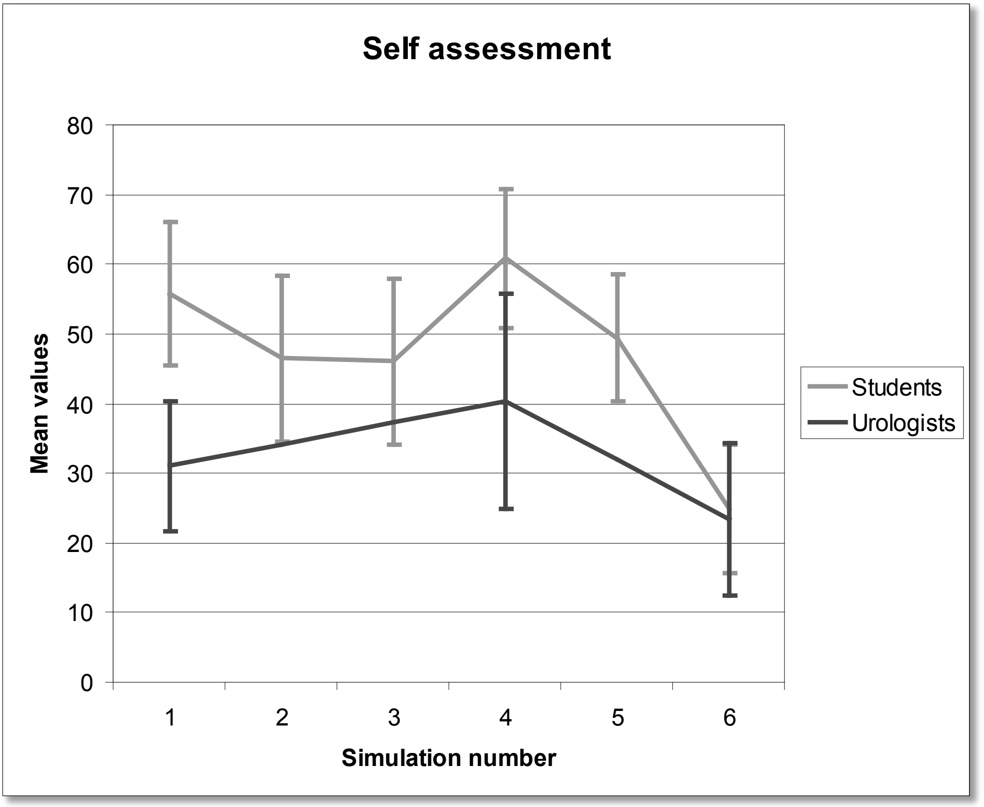

Figure 8. Another example of a learning curve. Data from the participants self assessments. A higher value represents an experience of a higher difficulty. Procedure 1, 2 and 6 were performed on the same easy difficulty case, 3 on a case with medium difficulty and 4 and 5 on a difficult case.

VR to OR

Altogether 24 participants from all over Sweden were included and divided randomly into two groups. There were no significant differences between the groups regarding sex (1/3 female), age (mean 33 years) or prior experience with transurethral procedures (mean 14 months of residency).

The course

The course evaluation showed that the participants believed they had learned most from the real TURP procedures and that patient safety was high. The simulated exercises were also rated high. During the course there was a significant improvement in knowledge concerning the instrumentation and the procedure (p=0.000 for both).

Simulator training

Figure 9 shows the results from the "criterion-procedures". The mean practise time before reaching the expert level was 198 minutes, not including the last validation procedure. The total mean practise time was 254 minutes distributed over 8.3 procedures of which 3.7 were criterion test-procedures.

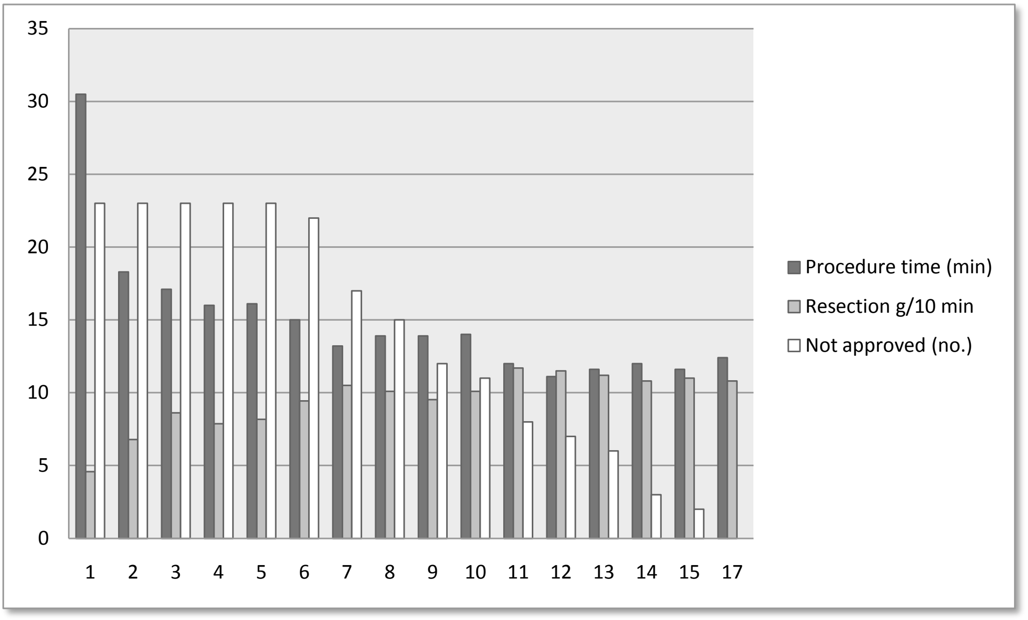

Figure 9. Data from the criterion based training in the simulator.

Peroperative assessment

The resection effectiveness measured as resection weight per minute was significantly lower (p=0.004). The number of aborted procedures due to poor skills/dexterity/judgment decreased from 30% to 0% (p=0.016), and the proportion of residents believed to be able to perform the procedure independent of a supervisor increased from less than 10% to about 75% (p=0.000). There were better scores on the checklist (p=0.000), global assessment (p=0.000) and self evaluation (p=0.000) (Figure 10).

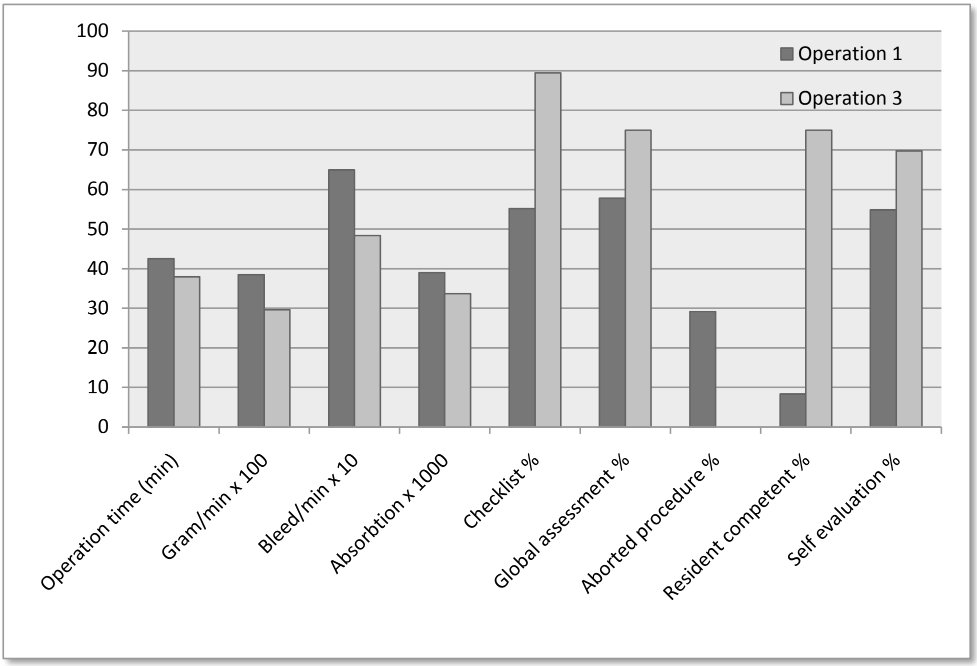

Figure 10. Data from the first and the last real operations on patients shows significant differences ($p < 0.001$, resected grams per minute: $p < 0.05$) except in operation time, bleed/min and absorption.

Inter-rater agreement

The rating was done by two surgically experienced urologists who analysed about 10 procedures together and then independent analyses of the same recordings and comparing the results until reaching an inter-rater agreement of 92%. If the calculation instead was based on how many times each action/error and progression score was noted for each recording, agreement reached 98%. This took into account the fact that the time-stamps were sometimes out of step between the raters by 1-2 minutes, depending on the video-player used to show the recordings.

Video assessment

When comparing the first and the last procedure there were a significant increase in the amount of autonomous procedure time (p=0,000), resection time (p=0,029) and a tendency to decreased haemostasis time and increased successful orientation (Figure 11).

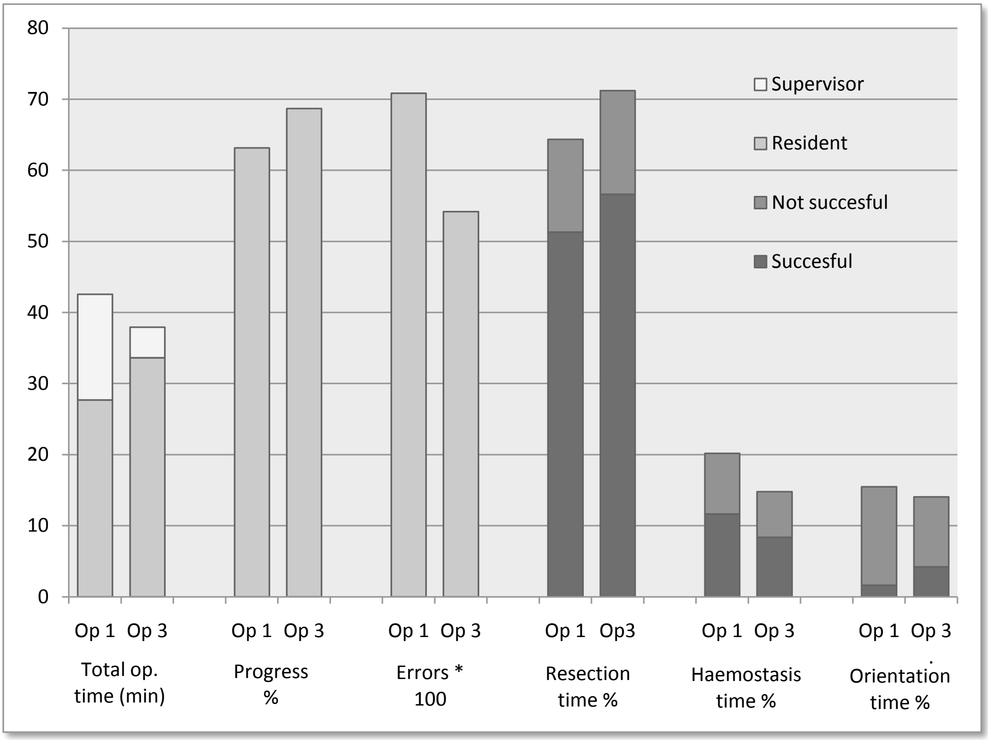

Figure 11. Data from the video assessment shows significant differences between the first and last real operation in resident operation time (p < 0.001) and resection time (p < 0.05).

Training transfer

Although it was not possible to measure any significant differences in single parameters (except for resident operation time which increased, p=0.025), there was a tendency for improved skills with simulation practise. When the number of participants who improved or showed no change in skills after simulation practise was compared with the results for procedures without simulation practise, there was a significant difference (p=0.021) indicating that simulation practise resulted in increased skills (Figure 12). Sixteen participants showed greater improvement after simulation practise compared to seven participants who showed greater improvement without practise. One participant could not be evaluated due to the exclusion of one patient.

Patient outcome

The 71 patients who underwent surgery had a follow-up time of 2.3 - 3.8 years. The mean prostatic size was 37.7 ml (19-67 ml) at preoperative TRUS-measure. There were no significant differences in prostatic size or in the order of operations between the groups. In three procedures, it was decided to perform an incision of the bladder neck due to small prostatic size. One patient was excluded due to an upper respiratory tract infection. The functional outcome was evaluated 6-12 months postoperatively. However, the patient response rate was low (46%), which makes these figures difficult to interpret (Table 15).

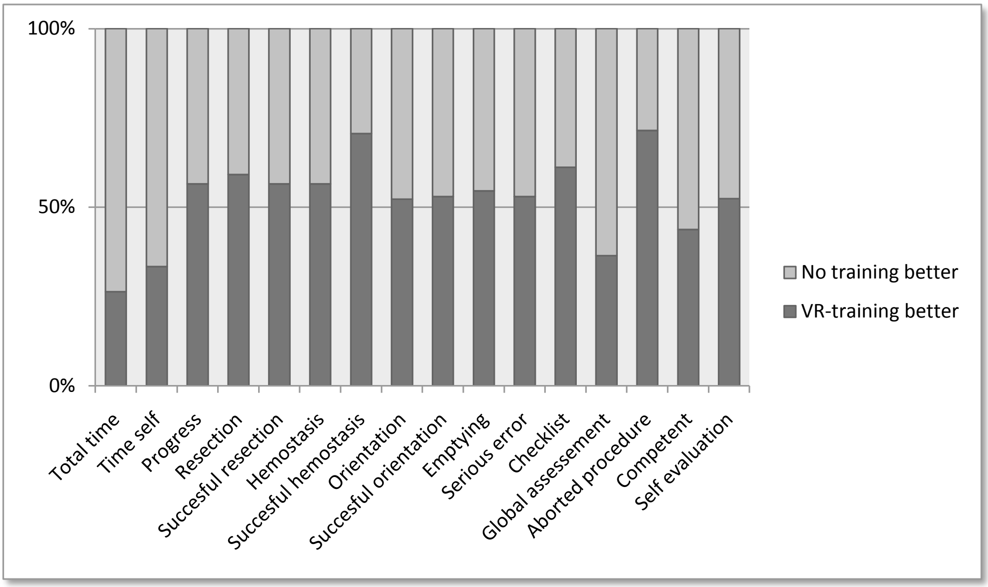

Figure 12. Data about transfer of skills from simulator training to operation room performance. The difference in performance between two consecutive operations without simulator training in between (No training better) is compared to two consecutive operations with simulator training in between (VR-training better).

The total mortality rate was 8 (11%), and three (4.2%) of these deaths could be linked to the surgery. Two of the patients had a CVI 26 and 51 days postoperatively and died 5 and 12 months later, respectively. The third case was more closely associated in time to the surgery. The TURP-procedure was performed without peroperative complications. In the postoperative period, the patient suffered from gross haematuria and angina pectoris. After ten days he had a myocardial infarction and was found dead in his home 78 days postoperatively. This gives a surgery-associated mortality rate of 1.4 - 4.2%. Two cases had surgery in the immediate postoperative period (misplaced suprapubic catheter, gross haematuria). Three patients were readmitted due to haematuria. Six patients underwent additional surgery (1 meatus-stenosis, 1 urethral stricture, 4 TURPs). Nine patients (12.7 %) suffered from incontinence six months postoperatively of whom eight had low-grade incontinence (score 1 of 5). One patient with complete incontinence was preoperatively diagnosed as having an overactive bladder. When analysing the video-recording it was not possible to identify any resection or damage to the area distal to the verumontanum (sphincter area).

	IPSS (0-35)		QoL (0-5)		Peak urine flow (ml/sec)		Incontinence %		ReTURP %	Mortality %
	Pre	Post	Pre	Post	Pre	Post	Pre	Post		
Follow-up	21.1	7.6	3.7	1.5	8.2	20.6	41	1-13	5.6	1.4-4,2
Reference	18-24	3-9	4-5	1-2	6-12	22-25	4-46	1-16	6-10	2-13

Table 15. Postoperative data were collected 6-12 months after the procedure except for the data regarding re-operations and mortality (2,3-3,8 years postoperative follow-up). References from large outcome studies[7,128,134-137].

Predictive validity

The study-group and set-up for the evaluation was the same as in Paper III. This study was designed to find out if there are any specific personal abilities or traits in the personality that together with the results from simulation practise can predict surgical performance. The Simulation learning score had a positive and significant correlation with the Video learning score (0.43, p=0.048) but not with OSATS or Operation learning score. Analysis of the theoretical assumptions association with the Operation learning score showed that a skilled performance in the OR was associated with a good Simulation learning score, a balanced attachment style, a good visual spatial memory and high levels of anticipatory worry, responsibility and goal directedness (Table 16).

Stepwise linear regression of Theoretical assumptions and Operation learning score (p = 0,031)	**Beta (p-value)**
Simulation learning score	0,90 (0,002)
Visual spatial memory	0,33 (0,129)
Poor impulse control	0,25 (0,295)
Balanced attachment style (RD3)	0,46 (0,053)
Anticipatory worry (HA1)	0,39 (0,052)
Responsibility (SD1)	0,36 (0,082)
Congruent second nature (SD5)	0,79 (0,004)

Table 16. The stepwise linear regression analysis of the theoretical assumptions and the Operation learning score.

Comparing the TCI-R scores with the population mean values[114] (Table 17) showed that the group had higher scores on the self-directedness, cooperativeness and reward dependence. On the subscales, the group scored somewhat higher on curiosity, attachment, social dependence, responsibility, purposefulness, resourcefulness, conscience and congruent second nature. The group had lower scores on the subscales shyness, and on all subscales related to self-transcendence.

TCI-R, Independent sample t-test Population mean = 50, SD = 10	Study group Mean (SD)	P-value
NS1 Curiosity	52,6 (5,5)	0,036
HA3 Shyness	45,4 (6,3)	0,002
RD Reward dependence	55,3 (7,3)	0,002
RD3 Attachment	56,8 (7,2)	0,000
RD4 Social Dependence	54,0 (8,7)	0,036
SD Self-directedness	57,0 (7,6)	0,000
SD1 Responsibility	56,2 (5,1)	0,000
SD2 Purposefulness	56,0 (8,6)	0,003
SD3 Resourcefulness	54,8 (4,6)	0,000
SD5 Congruent second nature	55,9 (8,4)	0,003
CO Co-operativness	53,4 (7,1)	0,037
CO5 Pure-hearted conscience	54,4 (8,1)	0,016
ST Self-transcendence	43,5 (10,2)	0,007
ST1 Self-forgetfulness	45,6 (8,4)	0,019
ST3 Spiritual acceptance	43,1 (10,6)	0,004

Table 17. Comparison of the participants and the general populations TCI-subscales mean values.

There were significant positive correlations between Operation learning score and Rey CFRT copy test (0,47, p=0,033), ToLdx in total execution time (0,50, p=0,020), problem-solving time (0,54, p= 0,012) and time violations (0,58, p=0,006). The psychometric scores compared with the population mean values (Table 18) show that the group had better executive planning abilities and better verbal working memory.

Psychometrics , Independent sample t-test	**Population Mean (SD)**	**Study group Mean (SD)**	**p-value**
ToLdx Total Move score	100 (10)	107,3 (11,5)	0,006
ToLdx Total problem-solving time	100 (10)	105,8 (9,5)	0,010
WAIS-III alphanumerical task	10 (3)	12,3 (2,7)	0,001

Table 18. Comparison of the participants and the general populations psychometric mean values.

Discussion

Face validity

One of the major difficulties was to obtain a level of realism that could be accepted by the experienced urologists without putting extreme demands on the hardware. Initially, the discussion was focused on e.g. making the appearance of bleeding more realistic but with time it became more important to create the same experience of difficulty as in the real situation and to be able to choose the right approach to solve the problem. With increasing simulator experience the tendency to focus on visual similarity on the real operation tends to decrease in favour of what can be performed on the simulator, for example practise different strategies or haemostasis control.

Content validity

The TURP procedure takes, in most cases, 30-60 minutes, depending on how difficult the case is and how experienced the urologist is. During the procedure a clear strategy between cutting tissue and coagulating must be achieved. If there is peroperative complications (i.e. great blood loss) a decision to abort the procedure has to be taken into consideration. Thus, besides practical skills there is also training in judgement. This makes the virtual resection simulator more close to the real life situation and to be a better tool for education.

Construct validity

We could show a significant difference in most of the variables and the clinically experienced group being more effective. The urologists tended to be more careful and more active during the procedure. The urologists showed a more aggressive technique, with more bleeding per minute (but with the same amount of total blood loss as the students) and a higher flow of irrigation fluid which has also been seen in other studies[13]. They also tended to rely more on their internalised view of the prostatic anatomy rather than letting the virtual anatomy guide the resection. They did not look as much at where the “capsule” was in the model (demonstrated by a bluish colour) but instead used their experience regarding the real anatomy. Compared to the students, they resected much more in the bladder neck and less in the area close to the *verumontanum*. The students used the information from the simulation environment to a greater extent and were careful to reach the depth of the capsule everywhere, without “perforations”. This could explain why there was no difference in resected volume and completeness.

When using the learning curve as a measurement of construct validity, inexperienced performers should have a positive learning curve (i.e. become more effective without higher risks) in every aspect which was true for the students in this study. Experienced performers should not show any significant improvement suggesting that they have already reached the plateau phase in their learning curve but even the urologists showed a positive learning curve in some parameters. This may be explained by an adaptation to the partly new environment that the simulation environment presents.

VR to OR

There was an improvement in skills during the course, and when the first operation was compared with the last, the majority of participants went from being unable to independently perform a simple TURP procedure to acquiring this ability. This was also reflected in a more skilled performance as assessed from the video recordings. The trainees tended to be more effective regarding successful actions and they spent less time attaining haemostasis, without any increase in blood loss or absorption of rinse solution. Resection effectiveness was lower regarding total weight resected and resection speed. This can be explained by the fact that the supervisor performed most of the resection in the first operation and in the last operation many of the participants performed the entire procedure independently. This study also shows that the surgical skills of urology residents increase more with than without preoperative simulator practise.

Predictive validity

The positive correlation between Rey CFRT copy test and learning scores indicate that visual spatial constructional ability is a better indicator of surgical skills then visual spatial memory[138]. The correlation between learning score and ToLdx (execution time, problem-solving time and time violations) could be interpreted as that it is better to take some extra time to solve the problem at hand. This is also in accordance with the importance of congruent second nature (SD5) and with anticipatory worry (HA1) in our model. Thus both psychometric tests and personality test are in concordance, which further emphasises the ability to plan and withhold impulses which could be in conflict with plans and to avoid errors.

When it comes to the ***personality character*** traits favourable in this experimental setting, above all high levels of *congruent second nature* (SD5) seems to be important. Congruent second nature refers to the subject's ability to control impulses. High scorers act in accordance with their long-term goals and values and rarely give in to temptation. During any operation there can be situations where the surgeon has to control impulses. This trait can thus be thought to be of importance, especially when learning new tasks where less control is possible. *Responsibility* (SD1) was a part of the model. High scorers have an internalised locus-of-control, which means that a person senses that problems are solvable due to own strength. These persons rarely blame others, but rather try to apply their own problem-solving skills to the problem. This trait was thought to *indirectly* influence performance in our model, but probably is a more important trait than first believed. For instance, the more a person feels “in charge” of the surgical procedure he is performing, the more flexible he is likely to be when confronted with possibly problematic events.

When it comes to the ***temperament profile***, *Anticipatory worry* and *Pessimism* (HA1) and a *Balanced Attachment style* (RD3) seemed to be favourable temperament traits. People who score high on HA1 manifest a distinctive behavioural tendency of being worriers that anticipate harm and failure, most pronounced in unfamiliar settings. This could probably be a useful trait to avoid errors when confronted with a relatively new and unfamiliar surgical procedure, as in our study. Attachment style refers to whether a person prefers intimacy to privacy. We proposed in our model that this trait indirectly could influence the learning situation and that that it would be favourable to be able to access both sides of this trait. Too much of either side could lead to problems in the learning situation, either being too sensitive to rejection and failure (high scorers), or being pronounced detached and disinterested in private relationships (low scorers).

Regarding the discussion about the personality of the surgeon, this study contributes with the finding that residents in urology seems to have personality traits that differs from the general population (although, still within the range of normality). The urology residents in this sample have a well developed character (high in SD and CO). Highly self-directed (SD) persons are described as effective, mature, reliable and goal-orientated. A person high in cooperativeness (CO) is often described as empathic, tolerant, and supportive and tries to cooperate with others as much as possible. The group had a higher than normal reward dependence which is a multifaceted higher order temperament trait. Individuals high in reward dependence tend to be tender-hearted, loving, warm, dedicated and sociable. Reward dependence was however not found to correlate with operation learning score, which suggest that this trait is perhaps more involved indirectly in learning situations, as we have proposed previously. It is likely that urologists in Sweden have similar traits since a profession tends to select apprentices that are similar to themselves[103]. These findings are in line with previous studies where 24 male and 15 female surgical residents[99] and 34 urology residents[139] were investigated by the use of the self-administered revised NEO Personality Inventory. The findings showed higher than average scores for extraversion (social, warm, talkative), openness (open-minded, willing to entertain new ideas) and conscientiousness (capable, determined, principled, purposeful, self-motivated).

Trait	Urology residents (n=24)	Anaesthetists (n=222)	Anaesthetists residents (n=54)	Physician (n=67)
Novelty seeking		-	-	-
Harm avoidance		+		
Reward dependence	+	-		-
Persistence		-	-	
Self directedness	+	+	+	+
Cooperativeness	+	+	+	+
Self transcendence	-	-	-	-

Table 19. Comparison of the findings in this study with the findings of Kluger and co-workers using TCI-125 on anaesthetists in New Zealand and physicians in Norway[140]. Mean values of the traits are compared to the general population. (+ = higher than population, - = lower than population, blank = no significant difference).

Methodological considerations

Face and content validity

In these studies it may have been better to use qualitative methods with interviews instead of questionnaires. Interviews may give a better insight of the problem and a clearer picture of what to focus on in the construction and refinement of the simulator and better definitions of difficulties experienced.

Construct validity

In this study a group of third-year medical students was tested. They had no prior experience of surgical procedures which clearly defines them as novices. The disadvantage with this is that they initially did not know anything about the procedure and the complications to avoid. They therefore attended a short course with information and instructions about the diagnosis, indications and procedure technique before they started the test. The motivation to perform the procedure probably differs between students prior to the choice of specialisation and the experienced urologists[141]. But when performing the test, the students were very focused and aware of the risks for the patient. The risk awareness was partly due to the “pre-operative” instructions, partly to the supervisor’s continuous instructions and partly due to the feedback from the ongoing procedure.

VR to OR

The choice to do a "cross-over" study in Papers III and IV partly depended on the size of the possible study group. The population of the Swedish urology residents that fulfilled the inclusion criteria is just slightly larger than the study group. In all practical aspects, the study group was equal to the specific population. If applying a randomised study design where the participants had been studied to a great extent at their home-clinics, the problems of bias, differences in tutoring and other practical aspects such as turnover time or impatience among the OR staff may have limited the possibilities to perform the procedures in a comparable way. With this study design we minimized confounding factors as much as possible. The training effect was also assessed as the individual's difference in performance and the participants were not compared with each other.

There were no significant differences in serious errors, which were very few in all, making the comparison difficult. The TURP procedure itself generates few serious errors, and the supervisors were also very aware and observant regarding possible errors and interrupted the participant or took over the procedure before the error was made. Nevertheless, there was a tendency for a successive reduction in errors during the week of the course.

It was not possible to measure more than a few significant differences when analysing single parameters. This may be explained by the small number of participants, a too short practise time in the simulator environment and/or too low level of difficulty of the criterion levels in the test-procedure. There are studies suggesting that task deconstruction is more beneficial than practising a full-task repeatedly[142-144] It may also be explained with the greater complexity when the trainee performs the entire procedure compared to when the supervisor takes over the procedure. When the supervisor takes over there is an accumulation of difficulties to take care of, and this is not reflected in the participant's score.

Predictive validity

The present study indicates that good coordination of visual spatial ability and psychomotor ability is important but there are also data that conflict with the theoretical assumptions that includes psychometric measurements and especially parameters from the RCFRT test. This could be explained by too small a data sample, insensitive tests or insensitive learning scores. The reason to choose this test-instrument was to give a better view of the visual spatial ability in combination with psychomotor ability (drawing a complex figure using the memorised picture) but it may have been better to use the mental rotation test that does not include the psychomotor aspect[79,83,145].

Bias

In the VR-to-OR and predictive validity studies the data favour VR-practise except for total procedure time, participant operation time, global assessment and the pass/fail-question. This may be explained by the fact that in the OR these data were very dependent on the supervisor. The procedures performed without VR-practise in between were done on two consecutive days, and both the supervisor and the participant may have felt more comfortable with the situation and with knowledge of the participant's skills fresh in mind.

The study design tried to compensate for bias by means of a strict validation structure and circumstances that were as homogeneous as possible. The participants had the same supervisor/evaluator during the entire course, which made the absolute value of the rating less important. The differences in skill were measured between the procedures and compared for each participant. The instrumentation, staff and OR were the same during the week and the only factor that differed substantially was the patient group.

Patient concerns

The procedure outcomes and patient safety were good compared with large outcome studies[134,135]. Different studies report a large variation regarding functional outcome after TURP[7,128,136,137]. The present data seem to be well within the reported range with the exception of peak urine flow, which is poorer than that reported in other studies. This may be explained by the low resection volume (less than 50% of the preoperative volume), or by the fact that there was little follow-up data (12%) for this figure. There may also be an over-representation of poor outcome since it is likely that unsatisfied patients contact the healthcare system more often than satisfied patients. The incontinence rate varies greatly between different studies (0-16%), which is probably due to different definitions of incontinence. In the present study there was an incontinence rate of 1.2 - 12.7% depending on how incontinence is defined. The follow-up shows no increased risks or poorer results for the patients compared to other studies.

Findings and implications

We conclude in the face and content validity studies that it is possible to develop a useful and safe simulator for training the skills of TURP.

A training course for learning the TURP procedure that includes simulator practise can significantly improve the participant's performance with the same patient outcome as when done in the usual clinical environment. Practising the TURP procedure in a VR-simulated environment improves performance even more and raises the level of skill and dexterity of inexperienced urology residents. To get the most out of a surgical simulator as a training tool, it should be a part of a larger curriculum that includes theoretical education concerning diagnostic tools, patient selection, treatment options and possible errors[146,147].

There are several advantages of measuring learning ability in a validated procedure simulator. One can practise the specific manual dexterity, the eye-hand (in this case also foot) coordination and also, to some extent, decision making. It gives a direct and objective measure of visual spatial and psychomotor abilities. A possible disadvantage is that a procedure simulator requires good knowledge of the procedure before the evaluation of surgical skills can be performed. When evaluating a beginner, it would be better to use a simulator with basic tasks, such as finding and grasping objects, moving objects etc. The important thing, however, is to see if the person has the ability to become a skilled surgeon, not if the person already has good knowledge of the specific procedure.

One way to reduce the risk for error is to evaluate the individual's need of education, both theoretically and practically. For the clinically active surgeon, the evaluation of necessary skills and knowledge needs to be an ongoing process, which should be initiated by both the surgeon and "the system". The period of residency is a completely different situation. In this case it is "the system" that should provide the possibilities to gain the necessary knowledge and skills, and also to do evaluations to ensure the resident reaches the specified goals before approval of responsibility and before starting to "practise" on patients[123,148].

TURP is still the most common urological procedure, although the number performed is diminishing due to pharmacological therapy.[3] The cost-benefit for treatment of BPH in developing countries favours surgery: a TURP procedure costs under \$1000[149], combined therapy (alpha-adrenergic and 5-alpha-reductase inhibitor) costs just over \$1000 per year, with a lifetime cost of 15 000 – 20 000 \$.[133] The use of pharmacological therapy may not be a problem in Sweden or in other industrialised countries, but it could be an important factor in developing countries where conservative or medical therapy may be impossible owing to poor compliance and lack of resources.[149-152] The development and evaluation of surgical simulators is not only of interest for industrialised countries, with growing demands for high quality surgery and ethical considerations, but also for the developing countries in need of surgical training.

The picture of the ideal surgeon is not based on any real knowledge of what the personal traits, abilities and skills should consist of. The employment procedure of a resident in urology/surgery is mainly based on formal criteria (such as scientific work, experience etc), some personal references and an interview[153]. The surgical profession puts a high demands on judgment and skills as in other professions such as pilots. The procedure to find a good pilot is much more complicated, involving repeated tests and evaluations of personality traits[154,155]. To reduce errors in surgery and to find the most suitable candidates for a profession in surgery, it may be in due time to introduce a more professional selection procedure[156-159].

Future studies and use of simulators

The higher order character dimensions *Self-directedness* and *Cooperativeness* did somewhat surprisingly not seem to be of importance in our study. This could be because all but one had normal/high levels in these dimensions. Perhaps it is not favourable as such to have normal/high scores but rather unfavourable to score low in these dimensions. This could only be investigated with a larger sample including a number of low scorers on SD and CO.

It would also be of interest to evaluate the importance of motivation in learning surgical skills.

To establish a predictive set of tools, the individuals in the present studies should be followed-up to see if the findings are consistent over time.

It may well be assumed that the balance between the ability to learn and psychomotor/ visual spatial abilities is important. An individual with good ability to learn may compensate for deficiencies in the other abilities, but requires more practise time than an individual with good abilities. It is however important to establish the lower limit for these abilities - when will it take an unreasonably long time to learn a task?

It would be of great interest to evaluate experienced and clinically skilled surgeons in the same manner as in this study, to get an idea of what one should look for when selecting surgical trainees.

We are planning to develop this TURP-simulator and to modify its content to create and validate a simulator for the training of TURB.

The use of simulators for the training of surgical skills is now a widespread reality. It is important to realise that a simulator is only one tool amongst others. Careful testing and selection of training goals is necessary before implementation into a surgical training curriculum.

Conclusions

We conclude that it is possible to construct, validate and use a full procedure, virtual reality, real-time simulator for training the skills of TURP.

We have established the requirements for and constructed a full-procedure, real-time, VR-simulator for the training of transurethral resection of the prostate (face and content validation).

Proof of construct validity was found in this simulator. It is concluded that the simulator could be used in the early learning phase, without significant risk for learning an incorrect methodological approach, to acquire a good strategy, and to learn to handle the instrumentation.

We found that a course for teaching the TURP procedure that includes training in a simulator can significantly improve the participant's performance.

Our findings suggest that practising the TURP procedure in a VR-simulated environment improves performance more than no practise at all. Practise in a simulator-based environment improves the skills and dexterity of urology residents when performing the procedure on patients, without increased risks for the patients.

Our findings suggest that simulation technology can be used to shorten the initial learning curve for residents in urology. The Visual Pelvic® system can be considered as clinically validated for this purpose

Swedish residents in urology seem to have personality traits that differ from the general population suggesting that they have a well-developed character (SD and CO) and higher than normal scores in reward dependence (RD). It is likely to believe that specialists in urology in Sweden have similar traits.

According to the findings of this study the optimal trainee has a good learning score in a simulated environment, is goal-directed, has a high level of impulse control, anticipates harmful events, take responsibility, has a balanced attachment style and has good visual spatial memory.

Svensk sammanfattning

Grunden i praktisk kirurgisk träning är den traditionella lärlingsmetoden vilken introducerades av William Halsted och har använts i över 100 år. Basen i detta lärande är träning i operationssalen där ST-läkaren under överinseende av en erfaren specialist gradvis lär sig operationsmetoder under operation av patienter. Den kontinuerliga utvecklingen av kirurgiska metoder och den allt ökande medicinska kunskapen har ändrat den kirurgiska vardagen jämfört med på Halsteds dagar. I dag ser vi en ökande specialisering med höga krav på specifik kunskap och kvalitet och en allt högre medvetenhet om medicinska risker och misstag. Undervisningssjukhusen har en hög omsättning av patienter med ofta komplicerade tillstånd och höga krav på resultat. Lägger man dessutom till de etiska aspekterna med att ”träna” kirurgi på patienter är det lätt att inse att möjligheterna att lära sig kirurgiska metoder på det traditionella sättet minskat.

När laparoskopisk kirurgi introducerades på bred front i slutet av 80-talet såg man ett ökat antal komplikationer vilket ledde till diskussioner om hur nya metoder skall introduceras och vilka krav detta ställer på kirurgen. Den laparoskopiska tekniken ger en tvådimensionell bild av en tredimensionell verklighet, man använder långa instrument som passerar bukväggen vilket medför att kirurgens rörelser blir av motsatt riktning inne i patienten. Detta ställer nya krav på kognitiva funktioner jämfört med den öppna tekniken. Samma begränsningar har dock gjort det möjligt att utveckla kirurgiska simulatorer för att träna laparoskopisk kirurgi utan inblandning av patienter. Denna utveckling tog fart i slutet på 90-talet och ett flertal kommersiella system finns idag. Innan introduktionen av nya verktyg och metoder är det viktigt att kontrollera deras innehåll, att validera funktionen och att försöka uppskatta nyttan av verktyget. Det finns idag rätt så goda bevis för nyttan av träning i flera av dessa simulatorer. Dessa system ger inte bara en möjlighet att träna kirurgiska ingrepp utan även en möjlighet att objektivt mäta specifika färdigheter.

Det finns idag ett fåtal specifika simulatorer för att träna urologiska ingrepp. Transuretral resektion av prostata (TURP) är det vanligaste urologiska ingreppet och görs för att minska besvären med lågt urinflöde vid godartad prostataförstoring. Ingreppet görs med endoskopisk teknik där man i ett instrument, som förs in till prostata via urinröret, kan skära ut prostatavävnad under ögats överinseende. Det är i teorin ett enkelt ingrepp men tar lång tid att lära sig (70-90 operationer) och kan under inlärningstiden medföra risker och försämrade resultat för patienten.

Urvalet av personer till högt specialiserade yrken såsom till exempel piloter och militärer görs genom att man testar specifika egenskaper och förmågor som man bedömer är viktiga för den kommande uppgiften. Detta är inget som regelmässigt görs i urvalet till medicinska specialiteter.

Syftet med denna studie var att ta reda på vilka delar av TURP-proceduren som uppfattades som svåra, att konstruera en realtids, virtual reality simulator för att träna detta, att utvärdera om simulatorn uppfyllde innehållskraven och om den gör nytta vid träning av detta ingrepp utan att öka patientrisken. Dessutom genomfördes tester av läkare under utbildning till urologer avseende personliga egenskaper och förmågor för att se om det finns specifika egenskaper/förmågor som är associerade med bättre resultat under kirurgisk träning.

Delarbete 1

En ”***face-validity***” studie genomfördes genom att ett frågeformulär sändes till 28 erfarna urologer efter att de fått studera en demonstrationsversion av TURP-simulatorn. Frågeformuläret innehöll frågor avseende deras upplevelse av vad de själva upplevt som svårt under inlärning av denna operation samt vad de upplevde att deras ”elever” har mest problem med. Baserat på de 17 besvarade enkäterna vidareutvecklades demonstrationsversionen av simulatorn. För att ytterligare kontrollera innehållet i simulatorn gjordes en ”***content validity***” test av nio erfarna urologer som efter att ha testopererat i simulatorn fyllde i ett nytt formulär. Baserat på detta resultat gjordes ytterligare förbättringar av simulatorns innehåll och därefter genomfördes en basal ”***construct validity***” test som involverade sju oerfarna läkarstudenter. Resultaten visade att simulatorn uppfyllde kraven på ”face-” och ”content-validity” då den innehöll de viktiga aspekterna av TURP operationen och att den möjliggjorde träning av detta. Läkarstudenterna uppvisade en positiv inlärningskurva under de sex operationerna som genomfördes vilket gav preliminära bevis även för simulatorns ”construct validity”.

Delarbete 2

Nio erfarna urologer genomförde tre operationer och elva oerfarna läkarstudenter genomförde sex operationer i simulatorn med syfte att bevisa simulatorns ”***construct validity***”. Data från operationerna sparades i filer, personernas beteende under operationerna noterades och personerna själva fick fylla i formulär efter varje operation för att beskriva upplevelsen av ingreppet. Resultatet visade att de oerfarna studenterna hade en tydlig inlärningskurva där de förbättrades avsevärt i de flesta mätvariabler. Urologerna uppvisade enstaka förbättringar i mätvärden sannolikt som tecken på en tillvänjning av simulatormiljön, som trots allt inte är helt lik den verkliga miljön. Det fanns också tydliga skillnader mellan grupperna där de erfarna urologerna uppvisade betydligt större effektivitet under operationen och med lägre ”risk” för den simulerade patienten.

Delarbete 3

För att testa den verkliga nyttan (***VR-to-OR***) med att träna i denna TURP-simulator skapades en kurs för att lära sig grunderna och behandlingsmetoderna vid benign prostataförstoring och med specifik inriktning på TURP. Målgruppen var ST-läkare i urologi under tidig fas av inlärning av TURP. De skulle ha en viss, men begränsad erfarenhet av transuretrala ingrepp. Kursen innehöll teoretiska moment avseende sjukdomen, diagnostiska metoder, behandlingsmetoder, urval av patienter till respektive behandlingsmetod samt riskerna och metoden vid TURP. De fick göra tre TURP operationer på patienter under noggrann övervakning av erfaren handledare som tog över operationen vid minsta misstanke på risk för patienten. Mellan två av operationerna fick de träna intensivt i TURP-simulatorn till dess att de uppfyllde kriterierna för godkänd. Dessa kriterier var baserade på medelvärdet av nio erfarna urologers resultat vid operation på samma simulerade patient. Resultaten från simulatorträningen sparades i datafiler som genererats av simuleringsprocessen. Bedömningen av operationer på patienter gjordes på två sätt. Den första delen gjordes av handledaren under operationen genom OSATS-metoden som består av en checklista med säkerhetspunkter som skall genomföras och av en sk ”global assessment form” som ger handledaren ett strukturerat stöd att bedöma kvalitén på operationens genomförande. Operationen videofilmades och en ”blindad” bedömning av ingreppet gjordes i efterhand. Kursen hölls vid tre tillfällen och totalt sett deltog 24 ST-läkare i urologi. Resultatet visar att kursen i sig gav signifikant förbättrad operationsteknik utan att öka vare sig peroperativa eller postoperativa risker för patienterna. Även den funktionella uppföljningen av patienternas vattenkastningsförmåga verkar ligga i nivå med vad som kan förväntas. Analysen visar också att man förbättrar sin färdighet vid TURP-operation genom träning i denna simulator.

Delarbete 4

St-läkare i urologi studerades avseende deras personlighet, visuospatiala förmåga, arbetsminne och exekutiva funktioner. Detta gjordes dels för att se om det fanns någon ***koppling mellan personliga egenskaper och kirurgisk förmåga*** och dels för att se om det finns personliga egenskaper som är specifika för urologer. Samma grupp av läkare som i delarbete 3, studerades med hjälp av:

- Personlighetstest - TCI (Temperament and Character Inventory)
- Test av visuospatiala förmågor - Rey complex figure and recognition trial
- Exekutiva förmågor - Tower of London (Drexel University)
- Arbetsminne – siffer- och bokstavs-repetitioner från WAIS-III
- Inlärningsdata från övningar i TURP-simulator

Dessa resultat jämfördes sedan med resultat från deras operationer på patienter samt med normalvärden från befolkningen.

Resultaten visar att dessa 24 ST-läkare i urologi har en välutvecklad personlighet jämfört med befolkningen, men inom normalvärden. De är mer effektiva, mogna, pålitliga, målinriktade, empatiska, toleranta, samarbetsvilliga samt är mer dedicerade och sociala. Samtidigt uppmättes lägre nivåer av ”andlighet” än hos normalbefolkningen. De uppvisar bättre färdigheter avseende exekutiv planering och verbalt arbetsminne. Vid regressionsanalys av dessa data mot operationsresultat fann vi att god kirurgisk inlärningsförmåga var kopplat till god inlärningsförmåga i TURP-simulatorn, bra visuospatialt minne, hög nivå av målinriktning, impulskontroll, förväntansångest och ansvarskännande samt en balanserad ”distans” till sin omgivning.

Diskussion och konklusion

Denna studie visar att det var möjligt att definiera och konstruera en simulator för att framgångsrikt träna metodiken vid TURP och att denna simulatorträning gav förbättrade operationsresultat. Denna simulator (PelvicVision®) kan anses vara ett bra verktyg för träning av transuretral resektion av prostata under tidig klinisk inlärningsfas och simulatorn kan med denna studie anses vara validerad för detta ändamål. Då vi med den kurs som skapades i samband med studien fick ansökningar att delta i kursen från i princip hela den svenska målgruppen ST-läkare i urologi belyses behovet av denna och liknande kurser. Metoden med expertbaserad nivå för godkänd kan förbättras och sannolikt kan mer träning i simulatormiljön ge ytterligare förbättringar av kirurgisk teknik vid riktiga operationer. Det förefaller rimligt att träning i en simulator till certifieringsnivå bör vara obligatoriskt innan träning på patienter tillåts. Våra studier av personlighet och personliga egenskaper bör i nuläget inte användas för selektion av blivande ST-läkare i urologi, men kan ligga till grund för fortsatta studier över vilka egenskaper som är lämpliga.

Acknowledgements

The path towards this thesis spans several years and many people have been involved in and contributed to the ideas and work presented. There are some people to whom I will always be in debt and to whom I wish to express my most sincere gratitude:

Joar Svanvik – my supervisor: for guiding and supporting me through this long and winding road. For always being helpful and giving good advice, gentle criticism, rapid response and gentle pressure through the years.

Hans Hjertberg – co-author and co-supervisor: As lecturer and mentor for the course participants he achieved a "super-hero" reputation when guiding them through their operations. Thanks for all the laughs and chat during endless hours of analysing video-taped operations.

Pether Carlsson – mentor for the course participants, guiding them through their first operations. For always taking care of things with a "no-problem" attitude.

Andreas Rousseau – my dear friend and co-author: for his important contribution, never ending enthusiasm and encouragement, and most of all for still being there - keep going.

Anders Spångberg – lecturer at the courses and roommate: for his patiently delivered and inexhaustible supply of good advice and answers to not always such good questions.

My colleagues and friends at the Department of Urology and especially the "Blue team" – ***Christer Ahlstrand, Bill Pettersson*** and ***Per Skoglund***: for encouragement and friendship and also for lecturing at the courses.

Johan Rosenqvist – head of the Department of Urology: for his encouragement and for providing the facilities necessary and being generous with time.

Pelle Larsson – former head of the Department of Urology: for believing in and encouraging the idea of having a course as a part of this thesis. Thanks for supporting me with your time and providing the necessary equipment.

The staff at the Department of Urology and especially ***Gunilla Weber, Roger Larsson*** and ***Carola Svahn***: for all their fantastic help and support.

Andreas Bengtsson – co-author: for essential and inspiring collaboration and for his important contribution as co-author.

Olle Eriksson and ***Kalle Wahlin***: for excellent statistical advice, fast answers and patience with sometimes rather stupid questions.

Leif Nord, Eva Skarman and the staff at Melerit: for their enthusiasm and friendship over many years of collaboration, and for being able to do so much with so little.

Sture Alexandersson: for his friendship, enthusiasm and never ending interest in making ideas come to life.

Clinicum - ***Stefan Redéen*** and ***Wolfram Antepohl***: for their generosity with time and to my new colleagues to whom I look forward to collaborate with in the future.

Henrik Kjölhede – co-author: for inspiring collaboration.

Peter Cox and ***Jane Wigertz*** – for immense and fast language support.

Andrea Capusan: for proof-reading and general support.

Per Källström: for valuable help with design and layout of the manuscript and cover.

Karl Storz – ***Anders Brifelt*** and ***Christer Waldton***: for generous support.

To all the medical students and residents in urology that participated in this study: without your participation, investing hours in strange tests and assessments, no results would have been obtained.

And last, but most certainly not least, to my family – ***Lina*** (especially for the cover-picture)***, Sanja*** and ***Oskar***, the essence of my life, and ***Christin,*** the love of my life: for your encouragement and support, for doing all the things I didn't have time for, and for being my guiding light.

The work constituting the basis of this thesis was supported by grants from the County Council of Östergötland and the Health Research Council of the South-East Region, Sweden.

References

1. Cameron JL: William Stewart Halsted. Our surgical heritage. Ann Surg. 225: 445-58, 1997.
2. Crothers IR, Gallagher AG, McClure N, James DT and McGuigan J: Experienced laparoscopic surgeons are automated to the "fulcrum effect": an ergonomic demonstration. Endoscopy. 31: 365-9, 1999.
3. Holtgrewe HL: Surgical management of benign prostatic hyperplasia in 2001--a pause for thought. J Urol. 166: 177, 2001.
4. de la Rosette JJ, Alivizatos G, Madersbacher S, Perachino M, Thomas D, Desgrandchamps F and de Wildt M: EAU Guidelines on benign prostatic hyperplasia (BPH). Eur Urol. 40: 256-63; discussion 264, 2001.
5. Wein AJ, Kavoussi LR, Novick AC, Partin AW and Peters CA: Campbell-Walsh Urology Online, in Wein AJ, Kavoussi LR, Novick AC, Partin AW andPeters CA, Elsevier, 2006, vol. 2010.
6. Hultén J and Hjertberg H: Studies on pressure volume response of the decentralized detrusor and its relation to the pressure in the pelvic veins. Scand J Urol Nephrol. suppl 82: 64-74, 1984.
7. Rassweiler J, Teber D, Kuntz R and Hofmann R: Complications of transurethral resection of the prostate (TURP)--incidence, management, and prevention. Eur Urol. 50: 969-79; discussion 980, 2006.
8. Mebust WK, Holtgrewe HL, Cockett AT and Peters PC: Transurethral prostatectomy: immediate and postoperative complications. A cooperative study of 13 participating institutions evaluating 3,885 patients. J Urol. 141: 243-7, 1989.
9. Ghanem AN and Ward JP: Osmotic and metabolic sequelae of volumetric overload in relation to the TUR syndrome. Br J Urol. 66: 71-8, 1990.
10. Hawary A, Mukhtar K, Sinclair A and Pearce I: Transurethral resection of the prostate syndrome: almost gone but not forgotten. J Endourol. 23: 2013-20, 2009.
11. Furuya S, Furuya R, Ogura H, Araki T and Arita T: [A study of 4,031 patients of transurethral resection of the prostate performed by one surgeon: learning curve, surgical results and postoperative complications]. Hinyokika Kiyo. 52: 609-14, 2006.
12. Holtgrewe HL, Mebust WK, Dowd JB, Cockett AT, Peters PC and Proctor C: Transurethral prostatectomy: practice aspects of the dominant operation in American urology. J Urol. 141: 248-53, 1989.
13. Cury J, Coelho RF, Bruschini H and Srougi M: Is the ability to perform transurethral resection of the prostate influenced by the surgeon's previous experience? Clinics (Sao Paulo). 63: 315-20, 2008.

14. Lindstedt E and Lundquist SB: [The development of transurethral surgery.]. Sydsven Medicinhist Sallsk Arsskr: 94-111, 1980.

15. Patel A, Quint RH and Fuchs GJ: Applications of electrical energy to the prostate: an evolutionary perspective. J Urol. 159: 1802-9, 1998.

16. Wanzel KR, Ward M and Reznick RK: Teaching the surgical craft: From selection to certification. Curr Probl Surg. 39: 573-659, 2002.

17. Fitts PM, Posner, M.I.: *Learning and skilled performance in human performance.* Belmont CA, Brock-Cole, 1967.

18. Schmidt RA: Schema theory of discrete motor skill learning. Psychological Review. 82: 225-260, 1975.

19. Ahlberg G, Kruuna O, Leijonmarck CE, Ovaska J, Rosseland A, Sandbu R, Stromberg C and Arvidsson D: Is the learning curve for laparoscopic fundoplication determined by the teacher or the pupil? Am J Surg. 189: 184-9, 2005.

20. Kruglikova I, Grantcharov TP, Drewes AM and Funch-Jensen P: The impact of constructive feedback on training in gastrointestinal endoscopy using high fidelity virtual reality simulation. A randomized controlled trial. Gut, 2009.

21. Ericsson KA and Lehmann AC: Expert and exceptional performance: Evidence of Maximal Adaptation to Task Constraints. Annual Review of Psychology. 47: 273-305, 1996.

22. Klingberg T, Fernell E, Olesen PJ, Johnson M, Gustafsson P, Dahlstrom K, Gillberg CG, Forssberg H and Westerberg H: Computerized training of working memory in children with ADHD--a randomized, controlled trial. J Am Acad Child Adolesc Psychiatry. 44: 177-86, 2005.

23. Ericsson KA: An expert-performance perspective of research on medical expertise: the study of clinical performance. Medical Education. 41: 1124-1130, 2007.

24. McGaghie WC, Issenberg SB, Petrusa ER and Scalese RJ: Effect of practice on standardised learning outcomes in simulation-based medical education. Med Educ. 40: 792-7, 2006.

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

26. Le CQ, Lightner DJ, VanderLei L, Segura JW and Gettman MT: The current role of medical simulation in american urological residency training programs: an assessment by program directors. J Urol. 177: 288-91, 2007.

27. Couves CM: A course in surgical technique for medical students. Can J Surg. 13: 31-2, 1970.

28. Lippert FG, 3rd, Spolek GA, Kirkpatrick GS, Briggs KA and Clawson DK: A psychomotor skills course for orthopaedic residents. J Med Educ. 50: 982-3, 1975.

29. Derossis AM, DaRosa DA, Dutta S and Dunnington GL: A ten-year analysis of surgical education research. Am J Surg. 180: 58-61, 2000.

30. Satava RM, Gallagher AG and Pellegrini CA: Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. J Am Coll Surg. 196: 933-7, 2003.

31. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D and Fried GM: Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. J Am Coll Surg. 198: 105-10, 2004.

32. Reznick RK: Teaching and testing technical skills. Am J Surg. 165: 358-61, 1993.

33. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C and Brown M: Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 84: 273-8, 1997.

34. Reznick R, Regehr G, MacRae H, Martin J and McCulloch W: Testing technical skill via an innovative "bench station" examination. Am J Surg. 173: 226-30, 1997.

35. Mackay S, Datta V, Chang A, Shah J, Kneebone R and Darzi A: Multiple Objective Measures of Skill (MOMS): a new approach to the assessment of technical ability in surgical trainees. Ann Surg. 238: 291-300, 2003.

36. Darzi A, Datta V and Mackay S: The challenge of objective assessment of surgical skill. Am J Surg. 181: 484-6, 2001.

37. Faulkner H, Regehr G, Martin J and Reznick R: Validation of an objective structured assessment of technical skill for surgical residents. Acad Med. 71: 1363-5, 1996.

38. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS and Jarvi KA: The educational impact of bench model fidelity on the acquisition of technical skill: the use of clinically relevant outcome measures. Ann Surg. 240: 374-81, 2004.

39. Grober ED, Hamstra SJ, Wanzel KR, Reznick RK, Matsumoto ED, Sidhu RS and Jarvi KA: Laboratory based training in urological microsurgery with bench model simulators: a randomized controlled trial evaluating the durability of technical skill. J Urol. 172: 378-81, 2004.

40. Harden RM, Stevenson M, Downie WW and Wilson GM: Assessment of clinical competence using objective structured examination. Br Med J. 1: 447-51, 1975.

41. Regehr G, MacRae H, Reznick RK and Szalay D: Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. Acad Med. 73: 993-7, 1998.

42. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Andersen DK and Satava RM: Analysis of errors in laparoscopic surgical procedures. Surg Endosc. 18: 592-5, 2004.

43. Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK and Satava RM: Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg. 236: 458-63; discussion 463-4, 2002.

44. Ahlberg G, Enochsson L, Gallagher AG, Hedman L, Hogman C, McClusky DA, 3rd, Ramel S, Smith CD and Arvidsson D: Proficiency-based virtual reality training

significantly reduces the error rate for residents during their first 10 laparoscopic cholecystectomies. Am J Surg. 193: 797-804, 2007.

45. Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, Reznick R and MacRae HM: Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. Surg Endosc. 18: 1800-4, 2004.

46. Messick S: Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist. 50: 741-749, 1995.

47. Gallagher AG, Ritter EM and Satava RM: Fundamental principles of validation, and reliability: rigorous science for the assessment of surgical education and training. Surg Endosc. 17: 1525-9, 2003.

48. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB and Jakimowicz JJ: Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc. 19: 1523-32, 2005.

49. Trindade JC, Lautenschlager MF and de Araujo CG: Endoscopic surgery: a new teaching method. J Urol. 126: 192, 1981.

50. Anastakis DJ, Regehr G, Reznick RK, Cusimano M, Murnaghan J, Brown M and Hutchison C: Assessment of technical skills transfer from the bench training model to the human model. Am J Surg. 177: 167-70, 1999.

51. Matsumoto ED, Hamstra SJ, Radomski SB and Cusimano MD: The effect of bench model fidelity on endourological skills: a randomized controlled study. J Urol. 167: 1243-7, 2002.

52. Aggarwal R, Moorthy K and Darzi A: Laparoscopic skills training and assessment. Br J Surg. 91: 1549-58, 2004.

53. Scott DJ, Bergen PC, Rege RV, Laycock R, Tesfay ST, Valentine RJ, Euhus DM, Jeyarajah DR, Thompson WM and Jones DB: Laparoscopic training on bench models: better and more cost effective than operating room experience? J Am Coll Surg. 191: 272-83, 2000.

54. Sutton C, McCloy R, Middlebrook A, Chater P, Wilson M and Stone R: MIST VR. A laparoscopic surgery procedures trainer and evaluator. Stud Health Technol Inform. 39: 598-607, 1997.

55. Okuda Y, Bryson EO, DeMaria S, Jr., Jacobson L, Quinones J, Shen B and Levine AI: The utility of simulation in medical education: what is the evidence? Mt Sinai J Med. 76: 330-43, 2009.

56. Tedesco MM, Pak JJ, Harris EJ, Jr., Krummel TM, Dalman RL and Lee JT: Simulation-based endovascular skills assessment: the future of credentialing? J Vasc Surg. 47: 1008-1; discussion 1014, 2008.

57. Gallagher AG and Cates CU: Approval of virtual reality training for carotid stenting: what this means for procedural-based medicine. JAMA. 292: 3024-6, 2004.

58. Schout BM, Hendrikx AJ, Scherpbier AJ and Bemelmans BL: Update on training models in endourology: a qualitative systematic review of the literature between January 1980 and April 2008. Eur Urol. 54: 1247-61, 2008.

59. Gurusamy KS, Aggarwal R, Palanivelu L and Davidson BR: Virtual reality training for surgical trainees in laparoscopic surgery. Cochrane Database Syst Rev: CD006575, 2009.

60. Seymour NE: VR to OR: a review of the evidence that virtual reality simulation improves operating room performance. World J Surg. 32: 182-8, 2008.

61. Ballaro A, Briggs T, Garcia-Montes F, MacDonald D, Emberton M and Mundy AR: A computer generated interactive transurethral prostatic resection simulator. J Urol. 162: 1633-5, 1999.

62. Kumar PV, Gomes MP, Davies BL and Timoney AG: A computer assisted surgical trainer for transurethral resection of the prostate. J Urol. 168: 2111-4, 2002.

63. Sweet R, Porter J, Oppenheimer P, Hendrickson D, Gupta A and Weghorst S: Simulation of bleeding in endoscopic procedures using virtual reality. J Endourol. 16: 451-5, 2002.

64. Sweet R, Kowalewski T, Oppenheimer P, Weghorst S and Satava R: Face, content and construct validity of the University of Washington virtual reality transurethral prostate resection trainer. J Urol. 172: 1953-7, 2004.

65. Kallstrom R, Hjertberg H, Kjolhede H and Svanvik J: Use of a virtual reality, real-time, simulation model for the training of urologists in transurethral resection of the prostate. Scand J Urol Nephrol. 39: 313-20, 2005.

66. Rashid HH, Kowalewski T, Oppenheimer P, Ooms A, Krieger JN and Sweet RM: The virtual reality transurethral prostatic resection trainer: evaluation of discriminate validity. J Urol. 177: 2283-6, 2007.

67. Padilla MA, Altamirano F, Arambula F and Marquez J: Mechatronic resectoscope emulator for a surgery simulation training system of the prostate. Conf Proc IEEE Eng Med Biol Soc. 2007: 1750-3, 2007.

68. Bach T, Geavlete B, Herrmann TR and Gross AJ: "Homemade" TUR-simulator for less than $40 U.S.? The "Tupper" experience. J Endourol. 23: 509-13, 2009.

69. Schout BM, Bemelmans BL, Martens EJ, Scherpbier AJ and Hendrikx AJ: How useful and realistic is the uro trainer for training transurethral prostate and bladder tumor resection procedures? J Urol. 181: 1297-303; discussion 1303, 2009.

70. Kallstrom R, Hjertberg H and Svanvik J: Construct Validity of a Full Procedure, Virtual Reality, Real-Time, Simulation Model for Training in Transurethral Resection of the Prostate. J Endourol. 24: 109-115, 2010.

71. Ritter FE, Schooler, L.J.: The learning curve: *International encyclopedia of the social and behavioral sciences.* Amsterdam, Pergamon, 2002, pp 8602-8605.

72. Schijven MP and Jakimowicz J: The learning curve on the Xitact LS 500 laparoscopy simulator: profiles of performance. Surg Endosc. 18: 121-7, 2004.

73. North M: The Hippocratic Oath, U.S. National library of medicine, 2009, vol. 2010.

74. Kohn LT, Corrigan JM and Donaldsson M: *To Err is Human: Building a safer health system.* Washington DC, Institute of Medicine, National Academy Press, 1999.

75. WHO: *World Alliance for Patient Safety, Forward programme.* 2005.

76. Satava RM: The nature of surgical error: a cautionary tale and a call to reason. Surg Endosc*.* 19: 1014-6, 2005.

77. Hedman L, Strom P, Andersson P, Kjellin A, Wredmark T and Fellander-Tsai L: High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. Surg Endosc*.* 20: 1275-80, 2006.

78. Schueneman AL, Pickleman J, Hesslein R and Freeark RJ: Neuropsychologic predictors of operative skill among general surgery residents. Surgery*.* 96: 288-95, 1984.

79. Hedman L, Klingberg T, Enochsson L, Kjellin A and Fellander-Tsai L: Visual working memory influences the performance in virtual image-guided surgical intervention. Surg Endosc*.* 21: 2044-50, 2007.

80. Lezak MD, Howieson, D.B., Loring, D.W., Hannay, H.J., Fischer, J.S.: *Neuropsychological Assessment.* New York, Oxford University Press, 2004.

81. Harris CJ, Herbert M and Steele RJ: Psychomotor skills of surgical trainees compared with those of different medical specialists. Br J Surg*.* 81: 382-3, 1994.

82. Francis NK, Hanna GB, Cresswell AB, Carter FJ and Cuschieri A: The performance of master surgeons on standard aptitude testing. Am J Surg*.* 182: 30-3, 2001.

83. Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED and Cusimano MD: Effect of visual-spatial ability on learning of spatially-complex surgical skills. Lancet*.* 359: 230-1, 2002.

84. McDougall EM, Corica FA, Chou DS, Uribe CA, Abdelshehid CS, Boker JR, Khonsari SS, Eichel L, Lee D, Finley DS *et al.*: Comparison of basic elements of human performance scores between urologists with various extents of experience. J Endourol*.* 20: 209-14, 2006.

85. Keehner MM, Tendick F, Meng MV, Anwar HP, Hegarty M, Stoller ML and Duh QY: Spatial ability, experience, and skill in laparoscopic surgery. Am J Surg*.* 188: 71-5, 2004.

86. Steele RJ, Walder C and Herbert M: Psychomotor testing and the ability to perform an anastomosis in junior surgical trainees. Br J Surg*.* 79: 1065-7, 1992.

87. Squire D, Giachino AA, Profitt AW and Heaney C: Objective comparison of manual dexterity in physicians and surgeons. Can J Surg*.* 32: 467-70, 1989.

88. Anastakis DJ, Hamstra SJ and Matsumoto ED: Visual-spatial abilities in surgical training. Am J Surg*.* 179: 469-71, 2000.

89. Carpenter PA, Just, M.: Spatial ability: An information processing pproach to psychometrics, in Sternberg RJ: *Advances in the psychology of human intelligence.* Hillsdale, NJ*,* Lawrence Erlbaum Associates*,* 1988*,* vol. 4, pp 221-253.

90. Ozer DJ: Personality, intelligence, and spatial visualization: correlates of mental rotations test performance. J Pers Soc Psychol*.* 53: 129-34, 1987.

91. Robert M and Chevrier E: Does men's advantage in mental rotation persist when real three-dimensional objects are either felt or seen? Mem Cognit. 31: 1136-45, 2003.

92. Salthouse TA and Mitchell DRD: *Effects of Age and Naturally Occurring Experience on Spatial Visualization Performance. [Miscellaneous Article].* Developmental Psychology September 1990;26(5):845-854.

93. Cronin-Golomb A, Braun, A. E. : Visuospatial dysfunction and problem solving in Parkinson's disease. Neuropsychology. 11, 1997.

94. Healey AN, Sevdalis N and Vincent CA: Measuring intra-operative interference from distraction and interruption observed in the operating theatre. Ergonomics. 49: 589-604, 2006.

95. Gazzaniga MS, Ivry, R.B., Mangun, G.R.: *Cognitive neuroscience - The biology of the mind.* New York, W. W. Norton & Company, Inc., 2002.

96. Hedman L, Klingberg T, Kjellin A, Wredmark T, Enochsson L and Fellander-Tsai L: Working memory and image guided surgical simulation. Stud Health Technol Inform. 119: 188-93, 2006.

97. Royall DR, Lauterbach, E. C., Cummings, J. L., Reeve, A., Rummans, T. A., Kaufer, D. I., LaFrance, W. C., Coffey, C. E.: Executice control function: A review of its promise and challenges for clinical research. The Journal of Neuropsychiatry & Clinical Neurosciences. 14: 377-405, 2002.

98. Hall JC, Ellis C and Hamdorf J: Surgeons and cognitive processes. Br J Surg. 90: 10-6, 2003.

99. McGreevy J and Wiebe D: A preliminary measurement of the surgical personality. Am J Surg. 184: 121-5, 2002.

100. Thomas JH: The surgical personality: fact or fiction. Am J Surg. 174: 573-7, 1997.

101. King RE and Flynn CF: Defining and measuring the "right stuff": neuropsychiatrically enhanced flight screening (N-EFS). Aviat Space Environ Med. 66: 951-6, 1995.

102. Maschuw K, Osei-Agyemang T, Weyers P, Danila R, Bin Dayne K, Rothmund M and Hassan I: The impact of self-belief on laparoscopic performance of novices and experienced surgeons. World J Surg. 32: 1911-6, 2008.

103. Greenburg AG, McClure DK and Penn NE: Personality traits of surgical house officers: faculty and resident views. Surgery. 92: 368-72, 1982.

104. Digman JM: Personality structure: Emergence of the five-factor model. Annual Review of Psychology. 41: 417, 1990.

105. Saulsman LM and Page AC: The five-factor model and personality disorder empirical literature: A meta-analytic review. Clin Psychol Rev. 23: 1055-85, 2004.

106. de la Rie SM, Duijsens IJ and Cloninger CR: Temperament, character, and personality disorders. J Pers Disord. 12: 362-72, 1998.

107. Ramanaiah NV, Rielage JK and Cheng Y: Cloninger's temperament and character inventory and the NEO Five-Factor Inventory. Psychol Rep. 90: 1059-63, 2002.

108. Cloninger CR, Svrakic DM and Przybeck TR: A psychobiological model of temperament and character. Arch Gen Psychiatry. 50: 975-90, 1993.

109. Cloninger CR: *The temperament and character inventory (TCI) : a guide to its development and use.* St. Louis, Mo., Center for Psychobiology of Personality, Washington University, 1994, pp 184 p.

110. Keltikangas-Jarvinen L, Ravaja N and Viikari J: Identifying Cloninger's temperament profiles as related to the early development of the metabolic cardiovascular syndrome in young men. Arterioscler Thromb Vasc Biol. 19: 1998-2006, 1999.

111. Svrakic NM, Svrakic DM and Cloninger CR: A general quantitative theory of personality development: Fundamentals of a self-organizing psychobiological complex. Dev. Psychoaffect. pathol. 8: 247-272, 1996.

112. Cloninger CR: A unified biosocial theory of personality and its role in the development of anxiety states. Psychiatr Dev. 4: 167-226, 1986.

113. DesCoteaux JG and Leclere H: Learning surgical technical skills. Can J Surg. 38: 33-8, 1995.

114. Brandstrom S, Schlette P, Przybeck TR, Lundberg M, Forsgren T, Sigvardsson S, Nylander PO, Nilsson LG, Cloninger RC and Adolfsson R: Swedish normative data on personality using the Temperament and Character Inventory. Compr Psychiatry. 39: 122-8, 1998.

115. Nicol DJ and Macfarlane-Dick D: Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. Studies in Higher Education. 31: 199 - 218, 2006.

116. Mahmood T and Darzi A: The learning curve for a colonoscopy simulator in the absence of any feedback: no feedback, no learning. Surg Endosc. 18: 1224-30, 2004.

117. Kaufman HH, Wiegand RL and Tunick RH: Teaching surgeons to operate--principles of psychomotor skills training. Acta Neurochir (Wien). 87: 1-7, 1987.

118. Säljö R: *Lärande i praktiken. Ett sociokulturellt perspektiv.* Stockholm, Prisma, 2000.

119. Macmillan AI and 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. 177: 274-7, 1999.

120. Grantcharov TP and Funch-Jensen P: Can everyone achieve proficiency with the laparoscopic technique? Learning curve patterns in technical skills acquisition. Am J Surg. 197: 447-9, 2009.

121. Francis NK, Hanna GB and Cuschieri A: The performance of master surgeons on the Advanced Dundee Endoscopic Psychomotor Tester: contrast validity study. Arch Surg. 137: 841-4, 2002.

122. Schlickum MK, Hedman L, Enochsson L, Kjellin A and Fellander-Tsai L: Systematic video game training in surgical novices improves performance in virtual reality

endoscopic surgical simulators: a prospective randomized study. World J Surg. 33: 2360-7, 2009.

123. Grantcharov TP, Bardram L, Funch-Jensen P and Rosenberg J: Impact of hand dominance, gender, and experience with computer games on performance in virtual reality laparoscopy. Surg Endosc. 17: 1082-5, 2003.

124. Ritter EM, McClusky DA, 3rd, Gallagher AG, Enochsson L and Smith CD: Perceptual, visuospatial, and psychomotor abilities correlate with duration of training required on a virtual-reality flexible endoscopy simulator. Am J Surg. 192: 379-84, 2006.

125. Cuschieri A, Francis N, Crosby J and Hanna GB: What do master surgeons think of surgical competence and revalidation? Am J Surg. 182: 110-6, 2001.

126. Gallagher AG, Smith CD, Bowers SP, Seymour NE, Pearson A, McNatt S, Hananel D and Satava RM: Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. J Am Coll Surg. 197: 479-88, 2003.

127. Darzi A, Smith S and Taffinder N: Assessing operative skill. Needs to become more objective. BMJ. 318: 887-8, 1999.

128. Marklund-Bau H, Edell-Gustafsson U and Spangberg A: Bothersome urinary symptoms and disease-specific quality of life in patients with benign prostatic obstruction. Scand J Urol Nephrol. 41: 32-41, 2007.

129. Meyers JE, Meyers, K. R. : *Rey Complex Figure Test and Recognition Trial – Professional Manual.* Lutz, Psychological Assessment Resources, Inc, 1995.

130. Culbertson WC, Zillmer, E. A. : *Tower of London Drexel University – second edition.* North Tonawanda, Multi-Health Systems, Inc, 2001.

131. Wechsler D: *Wechsler Adult Intelligence Scale – Third Edition.* San Antonio, The Psychological Corporation, 2003.

132. Dolmans VE, Schout BM, de Beer NA, Bemelmans BL, Scherpbier AJ and Hendrikx AJ: The virtual reality endourologic simulator is realistic and useful for educational purposes. J Endourol. 23: 1175-81, 2009.

133. Cury J, Coelho RF, Bruschini H and Srougi M: Is the ability to perform transurethral resection of the prostate influenced by the surgeon's previous experience? Clinics. 63: 315-20, 2008.

134. Madersbacher S, Lackner J, Brossner C, Rohlich M, Stancik I, Willinger M and Schatzl G: Reoperation, myocardial infarction and mortality after transurethral and open prostatectomy: a nation-wide, long-term analysis of 23,123 cases. Eur Urol. 47: 499-504, 2005.

135. Doll HA, Black NA, McPherson K, Flood AB, Williams GB and Smith JC: Mortality, morbidity and complications following transurethral resection of the prostate for benign prostatic hypertrophy. J Urol. 147: 1566-73, 1992.

136. Lourenco T, Armstrong N, N'Dow J, Nabi G, Deverill M, Pickard R, Vale L, MacLennan G, Fraser C, McClinton S *et al.*: Systematic review and economic modelling of effectiveness and cost utility of surgical treatments for men with

benign prostatic enlargement. Health Technol Assess. 12: iii, ix-x, 1-146, 169-515, 2008.

137. Melville A, Donovon J, Sheldon T and Peters T: Benign prostatic hyperplasia. Qual Health Care. 5: 111-9, 1996.

138. Stefanidis D, Korndorffer JR, Jr., Black FW, Dunne JB, Sierra R, Touchard CL, Rice DA, Markert RJ, Kastl PR and Scott DJ: Psychomotor testing predicts rate of skill acquisition for proficiency-based laparoscopic skills training. Surgery. 140: 252-62, 2006.

139. Eng MK, Macneily AE and Alden L: The urological personality: is it unique? Can J Urol. 11: 2401-6, 2004.

140. Kluger MT, Laidlaw TM, Kruger N and Harrison MJ: Personality traits of anaesthetists and physicians: an evaluation using the Cloninger Temperament and Character Inventory (TCI-125). Anaesthesia. 54: 926-35, 1999.

141. Kaplan AG, Kolla SB, Gamboa AJ, Box GN, Louie MK, Andrade L, Santos RT, Gan JM, Moskowitz RM, Shell C *et al.*: Preliminary evaluation of a genitourinary skills training curriculum for medical students. J Urol. 182: 668-73, 2009.

142. Ghahramani Z and Wolpert DM: Modular decomposition in visuomotor learning. Nature. 386: 392-5, 1997.

143. Kishore TA, Beddingfield R, Holden T, Shen Y, Reihsen T and Sweet RM: Task deconstruction facilitates acquisition of transurethral resection of prostate skills on a virtual reality trainer. J Endourol. 23: 665-8, 2009.

144. Gershman SJ, Pesaran B and Daw ND: Human reinforcement learning subdivides structured action spaces by learning effector-specific values. J Neurosci. 29: 13524-31, 2009.

145. Wanzel KR, Hamstra SJ, Caminiti MF, Anastakis DJ, Grober ED and Reznick RK: Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. Surgery. 134: 750-7, 2003.

146. Kohls-Gatzoulis JA, Regehr G and Hutchison C: Teaching cognitive skills improves learning in surgical skills courses: a blinded, prospective, randomized study. Can J Surg. 47: 277-83, 2004.

147. Van Herzeele I, Aggarwal R, Neequaye S, Darzi A, Vermassen F and Cheshire NJ: Cognitive training improves clinically relevant outcomes during simulated endovascular procedures. J Vasc Surg. 48: 1223-30, 1230 e1, 2008.

148. Grantcharov TP and Reznick RK: Training tomorrow's surgeons: what are we looking for and how can we achieve it? ANZ J Surg. 79: 104-7, 2009.

149. Alhasan SU, Aji SA, Mohammed AZ and Malami S: Transurethral resection of the prostate in Northern Nigeria, problems and prospects. BMC Urol. 8: 18, 2008.

150. Riviello R and Ozgediz D: International Medical Graduates and the Global Surgical Workforce: The Perspective from the Other Side. Journal of the American College of Surgeons. 207: 143-144, 2008.

151. Ozgediz D, Kijjambu S, Galukande M, Dubowitz G, Mabweijano J, Mijumbi C, Cherian M, Kaggwa S and Luboga S: Africa's neglected surgical workforce crisis. Lancet. 371: 627-8, 2008.

152. Laxminarayan R, Ashford L.: Using Evidence About "Best Buys" to Advance Global Health, Disease Control Priorities Project, 2008, vol. 2008.

153. Kron IL, Kaiser DL, Nolan SP, Rudolf LE, Muller WH, Jr. and Jones RS: Can success in the surgical residency be predicted from preresidency evaluation? Ann Surg. 202: 694-5, 1985.

154. McGreevy JM: The aviation paradigm and surgical education. J Am Coll Surg. 201: 110-7, 2005.

155. Carretta TR: U.S. Air Force pilot selection and training methods. Aviat Space Environ Med. 71: 950-6, 2000.

Appendix

Patient nr	1 - easy	2 -medium	3 -difficult
Distance between lateral lobes (mm)	2	7	2
Prostatic angle (degrees)	30	30	40
Prostatic radius (mm)	15	19	20
Prostatic volume (ml)	15	29	37
Fog factor	0,5-0,6	0,5-0,6	0,75-0,85
Tissue stiffness factor	3	3	5
Vessel density (nr/degree/mm)	0,004-0,0004	0,004-0,0004	0,008-0,0008
Distance to start of arterial vessels (mm)	0,3-0,9	0,3-0,9	0,5-0,8
Distance to start of venous vessels (mm)	0,1-0,5	0,1-0,5	0,3-0,6
Arterial pressure (mm Hg)	85	85	85
Venous pressure (mm Hg)	15	15	10
Arterial vessel area (mm^2)	0,05-0,1	0,05-0,1	0,05-0,2
Venous vessel area (mm^2)	0,05-0,1	0,05-0,1	0,05-0,3
Bladder volume (ml)	500	500	700
Tissue force factor	50	50	0
Risk factors	Hypertension, diabetes	Hypertension, angina pectoris, diabetes	Angina pectoris, by-pass surgery, heart failure

Appendix 1. The simulated patient cases used during the content and construct validity tests.

What were the three most difficult parts of the procedure?

……………………………………………………………………………………………..

……………………………………………………………………………………………..

……………………………………………………………………………………………..

……………………………………………………………………………………………..

……………………………………………………………………………………………..

……………………………………………………………………………………………..

How did you experience the following?

	Easy	Difficult
General impression	| -- |	
Vision, orientation	| -- |	
Handling instruments	| -- |	
Cutting	| -- |	
Coagulating	| -- |	
Visualising and understanding the anatomy	| -- |	
Using a strategy	| -- |	
Control of patient status	| -- |	
Difference from previous procedure	| -- |	

	Good	Bad
Final result	| -- |	

Appendix 2. The study participant self-evaluation questionnaire.

Total time of the procedure
Inactive time, i.e. when not cutting, coagulating or emptying the bladder
Time with high intravesical pressure
Time spent cutting and coagulating tissue
Time pressing the cutting or coagulation foot pedal without contact with the tissue
Time spent cutting or coagulating tissue during good vision of the cutting area and time with very poor or no vision
Total volume of resected tissue and volume of adenoma remaining
Mean and maximum volumes of cut-away chips
Volume of blood loss
Number of arteries still bleeding when the trainee has finished
Volume of irrigation fluid used and absorbed by the patient and number of times the bladder was emptied
Total distance travelled by the sling and resectoscope tip

Appendix 3. The parameters measured and saved during the simulated resection of the prostate.

	Yes	No
Check patient id		
Check operation data (prostatic size, risks, antibiotics, etc.)		
Check instruments		
Check foot pedals		
Check diathermy apparatus		
Check focus on video screen		
Check white balance		
Correction of working position		
Safe introduction of resectoscope		
Visualisation of anatomy		
Active control of rinse solution		
Empties rinse solution regularly		
Visualisation of the capsule		
Inverted grip on resectoscope when needed		
Control of patient status		
Realises when to end or when to ask supervisor to take over		
Empties bladder of resection chips		
Use of an appropriate KAD		
Use of rinse solution when needed		
Informs the patient about the results		
Informs the staff about the results and the postoperative plan		

Appendix 4. The checklist used in the OSATS-evaluation during the construct and VR-to-OR studies.

Score	1	2	3	4	5
Respect for tissue	Resectoscope or loop frequently forced against tissue. Not careful.		Resectoscope or loop sometimes forced against tissue. Mostly careful.		No unnecessary forces against tissue. Careful.
Time and movements	Many unnecessary movements.		Some unnecessary movements, mostly time effective.		No unnecessary movements, optimal use of time
Eye-hand coordination	Glances back and forth between monitor and hand/resectoscope most of the time.		Glances back and forth between monitor and hand/resectoscope some of the time.		Almost never glances back and forth between monitor and hand/resectoscope.
Foot pedals	”Searches” for the foot pedals a great deal of the time.		”Searches” for the foot pedals some of the time.		Almost never “searches” for the foot pedals.
Videoscope	Frequent problems with aiming the instrument. Poor visualization.		Mostly correct aim and visualization.		Always good aim and optimal visualization.
Resection	Frequently cuts too large or too small chips				Frequently cuts good-sized chips.
Strategy	Frequently interrupts procedure and needs supervision.		Ability to plan ahead and relatively good progress.		Obvious planning ahead from start to finish.
Tempo	Progresses at very slow or too fast a tempo.		Progresses at a somewhat low or fast tempo but without obvious risks for damage		Progresses at a good tempo without risks
Use of assistance	Does not communicate with the staff		Sometimes communicates with the staff		Strategic communication with the staff during the entire procedure
Stress level	Clearly stressed		Sometimes stressed		Calm, methodical
Supervision	During most of the procedure		During parts of the procedure		Almost none
Communication with supervisor	Does not ask for advice despite obvious need		Sometimes asks for advice but not for enough advice		No need for advice or asks for advice when necessary
Knowledge about the procedure	Poor knowledge, need for specific instructions during most of the procedure		Knowledge about the important aspects of the procedure		Good knowledge about all aspects of the procedure
Final result	Not near completion		Requires completion by the supervisor		Completed

☐ **Supervisor takes over the procedure due to inappropriate performance by the resident**

Do you believe the resident is competent enough to independently perform this procedure on a small to moderately enlarged prostate?

Yes No

Appendix 5. The global assessment form used in the OSATS-evaluation during the construct and VR-to-OR studies, modified from Seymour et al[42].

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	…
Lack of progress															
Attending takeover															
Damage (ca, s, b, u, co)															
Resection															
Haemostasis															
Emptying															
Orientation															
Technical issues															
Progress															

Appendix 6. The video assessment form used in the VR-to-OR study.