Doctor of Medicine (MD)

The validation and application of a novel colonic polypectomy trainer

---

The WIMAT colonoscopy suitcase

By

James Ansell

March 2014

Summary

Background and Aims

The WIMAT colonoscopy suitcase is an ex-vivo, porcine, polypectomy simulator. This has been developed in response to the increasing demand for polypectomy training following the introduction of the National Bowel Cancer Screening Programme. The aims of this thesis are to establish if the simulator is a valid form of polypectomy skills training and to identify if this model can be used to develop objective parameters for polypectomy assessment.

Materials and Methods

A series of clinical trials were systematically conducted to test the validity of the WIMAT colonoscopy suitcase. This included evaluating its content, construct and concurrent validity and conducting a skills transfer study comparing the WIMAT colonoscopy suitcase with a virtual reality simulator. Objective assessment parameters were examined by measuring the accuracy of self-assessment and using video coding software to analyse the hand movements performed during simulated polypectomy tasks.

Results

Content validity was demonstrated by experts who scored the model's anatomical, mechanical and visual realism favourably across multiple parameters ($p=<0.01$). Construct and concurrent validity were confirmed by participants performing simulated polypectomy in accordance with their "real-life" level of expertise ($p=<0.01$). Skills transfer to the clinical setting was demonstrated in a pilot randomised controlled study. Self-assessment following simulated polypectomy is inaccurate as experts tend to overestimate ability whereas novices underestimate ability ($p=>0.05$). The ratio of rotational hand movements to endoscopic tip angulation (RoTA) was significantly different when comparing novices to experts ($p=<0.05$).

Discussion

The WIMAT colonoscopy suitcase is a valid form of polypectomy skills training. The simulator can be used to address the increasing demand for training in this procedure. Further work is needed to assess the reliability of the RoTA score at different stages of the polypectomy procedure before it is used as an assessment tool.

Acknowledgements

Many people have contributed to the work described in this thesis and their support is gratefully received. Firstly, I would like to express my gratitude to my principal supervisor, Mr Jared Torkington for his full support, expert advice and encouragement over the last two years. I would also like to thank my academic supervisors, Dr Neil Warren and Professor Keith Harding who have been invaluable in generating ideas, appraising work and guiding me through the development of this research. I am also grateful to Professor Patricia Price who kindly shared her time to advise on the statistical analysis of these results. I would also like to thank Mr Konstantinos Arnaoutakis and Mr Stuart Goddard for their methodological advice and technical support. Drs James Horwood, Joanna Hurley, Raji Ramaraj and Medical Students' Chantelle Rizan and James Glasbey all assisted with data collection. Their contributions are all grateful acknowledged.

This work has been funded by a generous grant from the Royal College *of* Surgeons *of* England Research Fellowship. I am very appreciative of Mr Martyn Coomer and his team for their help and support. I would especially like to thank Mr Brian Rees who has been a valued role model and mentor to me throughout my surgical career.

Finally, I would like to thank my wife, Tania and my family for their unconditional love and support during the last two years. I would not have been able to complete this thesis without their continuous encouragement.

List of abbreviations

A	Advanced
AD	Advanced skills in colonoscopy
ABP	Animal By-Products
BA	Basic Skills in Colonoscopy
BE	Bench Model
CAE	Canadian Aviation Electronics
CI	Confidence Interval
CO	Cohort
CONSORT	Consolidated Standards Of Reporting Trials
DOPS	Directly Observed Procedural Skills
DOPyS	Direct Observation of Polypectomy Skills
E	Expert
EGT	Eye Gaze Technology
ERIC	Education Resources Information Centre
EMR	Endoscopic Mucosal Resection
ESD	Endoscopic Submucosal Dissection
EV	Ex-vivo
EUS	Endoscopic Ultrasound
FOB	Faecal Occult Blood
FN	Foundation Skills
GI	Gastro Intestinal
I	Intermediate
ICF	Informed Consent Form
ID	Identification
IN	Intermediate skills in colonoscopy
IQR	Inter-Quartile Range
IRGUS	Image Registered Gastroscopic Ultrasound System
ISRCTN	International Standard Randomised Controlled Trial Number
JAG	Joint Advisory Group on Gastro Intestinal Endoscopy
JETS	JAG Endoscopy Training System
LA	Live Animal
MeSH	Medical Subject Headings
MINS	Minutes
MOD	Module
N	Novice
NBCSP	National Bowel Cancer Screening Programme

NHS	National Health Service
NOTES	Natural Orifice Transluminal Endoscopic Surgery
OGD	Oesophago-Gastro-Duodenoscopy
PASW	Predictive Analytics SoftWare
PDP	Personal Development Plan
PEG	Percutaneously Endoscopic Gastroscopy
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analyses
PY	Polypectomy training course
RAF-c	Rotterdam Assessment Form for colonoscopy
RCT	Randomised Controlled Trial
REC	Research Ethics Committee
RoTA	Rotation To Angulation
SE	Standard Error
SPSS	Statistical Package for the Social Sciences
ST	Specialty Trainee
TEM	Transanal Endoscopic Microsurgery
TEO	Transanal Endoscopic Operation
VGP	Visual Gaze Pattern
VR	Virtual Reality
WIMAT	Welsh Institute for Minimal Access Therapy

"Better is possible. It does not take genius. It takes diligence. It takes moral clarity. It takes ingenuity. And above all, it takes a willingness to try."

Atul Gawande, 2007

Contents

PART 2: ASSESSMENT USING THE WIMAT COLONOSCOPY SUITCASE

Part 1:

Validation of the WIMAT colonoscopy suitcase

Chapter 1: Introduction

The diagnosis, management and treatment of adenomatous polyps

1.1 Colorectal cancer and adenomatous polyps

a) Epidemiology, aetiology and pathogenesis

Colorectal carcinoma is the second most common cancer in the UK (Statistics, 2011). In 2009 there were 18,538 new cases for men and 15,066 for women (Statistics, 2011). This is equivalent to an incidence rate of 57 new cases per 100,000 men and 38 per 100,000 women (Statistics, 2011). It is the third most common cause of cancer death, after lung and prostate cancer in men, and lung and breast cancer in women (Statistics, 2011). Diet and Western lifestyle are associated with the development of colorectal cancer but no specific food or environmental agent has been identified as a true causative factor (Ponz de Leon, 1996, Potter, 1999). Most colorectal cancers are however, thought to arise from adenomatous polyps (adenomas) (Riley, 2008).

The adenoma-carcinoma sequence describes the stepwise progression from normal tissue to dysplastic colonic epithelium and subsequent colorectal carcinoma. This is associated with multiple clonally selected genetic alterations (**Figure 1**) (Fearon and Vogelstein, 1990, Leslie et al., 2002). Adenomas are the dysplastic precursor lesions described in this model. These are non-invasive tumours of epithelial cells arising from the mucosa with the potential to become malignant (Bujanda et al., 2010). Adenomas are found in up to 30% of patients over 60 years of age in North America and Europe (Atkin et al., 1992, Papatheodoridis et al., 1998).

b) Classification of adenomatous polyps

Adenomas can be classified on a morphological or histological basis. The *Paris* classification (**Figure 2**) describes the morphological appearance of polyps as either polypoid (pedunculated or sessile) or nonpolypoid (flat or ulcerated) (Workshop, 2003). Alternatively,

the *Pit Pattern* classification groups each polyp according to its surface appearance **(Table 1 and Figure 3)** (Kudo et al., 1996). The morphology of a polyp can be used as a guide to its potential malignancy (Lieberman et al., 2008). The probability of high grade dysplasia and of carcinomatous transformation increases with polyp size (>1cm), depression, irregular contours and deformity. Short and immobile stalks and a polyp that poorly elevates after submucosal injection are further signs of underlying malignancy (Bujanda et al., 2010).

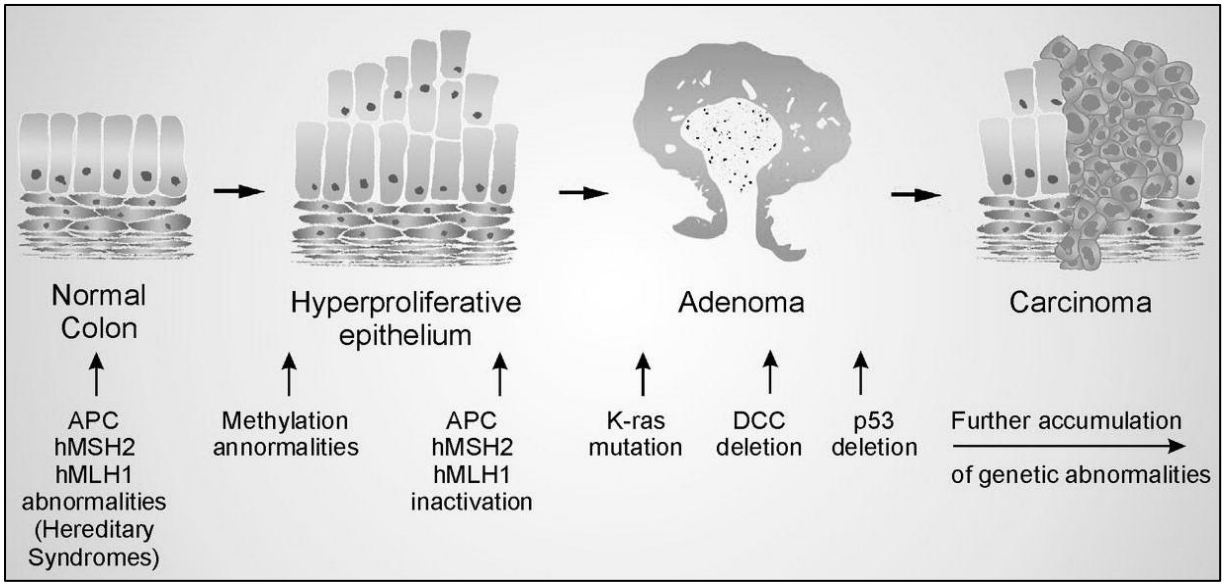


Figure 1: Adenoma-Carcinoma Model of colorectal tumorgenesis (Fearon and Vogelstein, 1990).

Adenomas can be classified histologically as tubular (<20% villous architecture), villous (80% villous architecture) or tubulovillous (O'Brien et al., 1990). Approximately 87% of adenomas are tubular, 8% tubulovillous and 5% villous (O'Brien et al., 1990). The Haggitt classification (**Figure 4**) assigns levels of malignant invasion to each polyp (Haggitt et al., 1985). Level 1 describes adenocarcinoma limited to the polyp head; level 2 includes neck involvement; level 3 corresponds to adenocarcinoma in the stalk and level 4 to invasion into the submucosa (Bujanda et al., 2010).

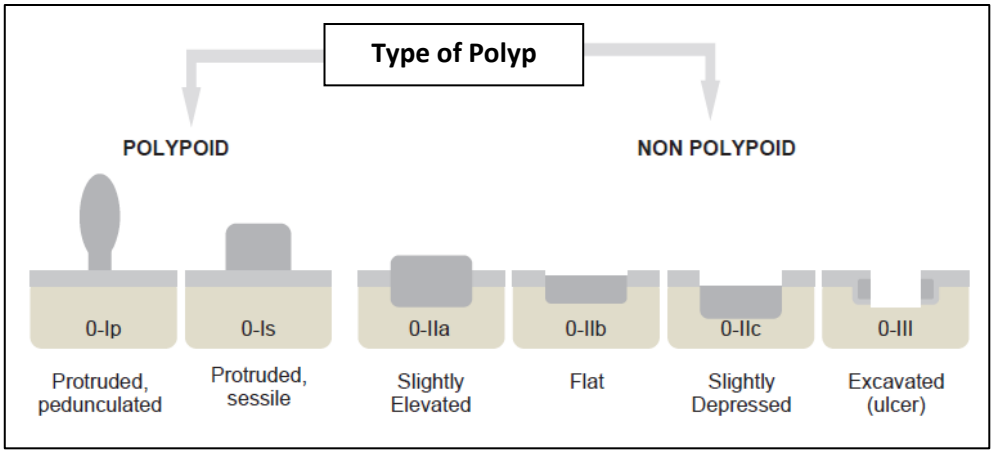


Figure 2: The Paris morphological classification of neoplastic lesions (Workshop, 2003)

Type of lesion	Description
Type 1	Round pits
Type 2	Stellar or papillary pits
Type 3L	Large tubular or round pits
Type 3S	Small tubular or round pits
Type 4	Branch-like or gyrus-like pits
Type 5	Non-structural pits

Table 1: Kudo pit pattern classification (Kudo et al., 1996). Tumorous lesions have either Type 3L, 3S, 4 and or 5 pit patterns (Bujanda et al., 2010)

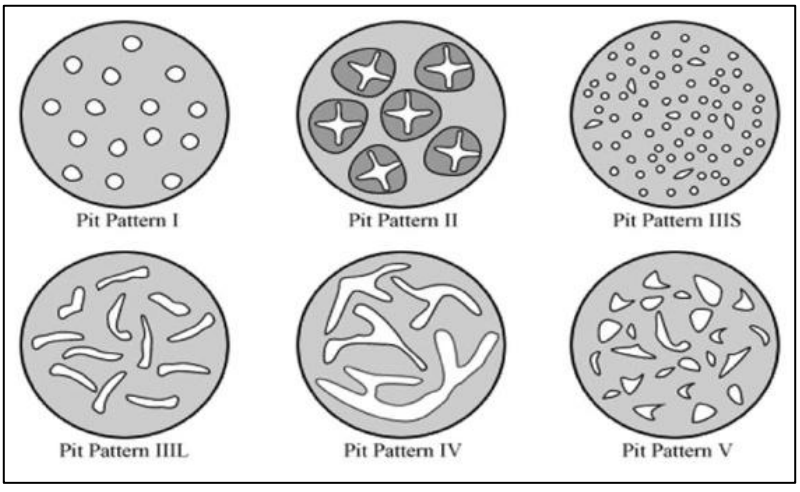


Figure 3: Kudo pit pattern classification in diagrammatic form (Aabakken, 2010)

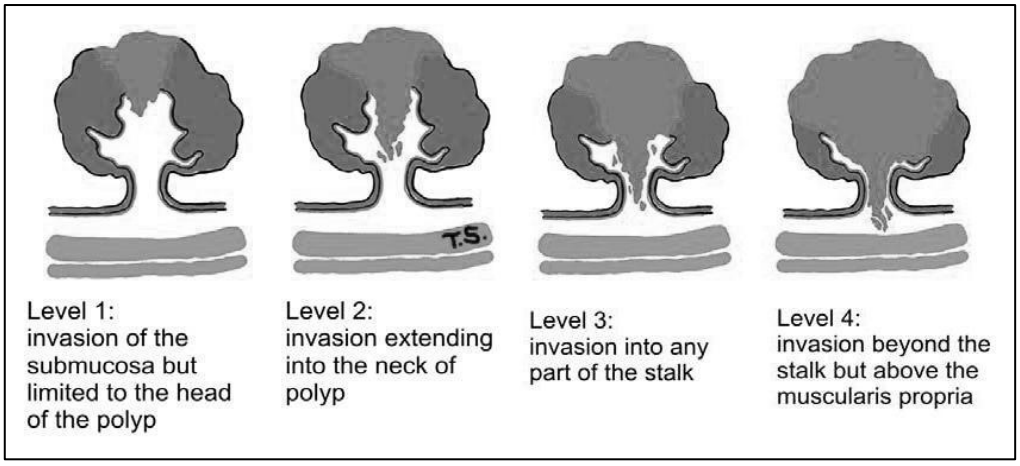


Figure 4: The Haggitt Classification (Haggitt et al., 1985)

c) Management of adenomatous polyps

It is current practice to remove polyps when detected, search the colon for additional lesions, and arrange for long-term follow-up of the subject (Winawer et al., 1990). This is based on the concept that adenomatous polyps are the precursor of colorectal cancer and that removing them will prevent malignancy from occurring (Muto et al., 1975).

d) Colonic polyp surveillance in the UK

Following the removal of adenomas, 30-35% of patients will have further polyps detected at 3-4 years (van Stolk et al., 1998, Neugut et al., 1985). This has led to a policy of endoscopic surveillance for all adenoma bearers. Although only 3% of patients with colonic adenomas will go on to suffer from colorectal cancer, there are no reliable criteria available that can predict adenoma progression or recurrence (Leslie et al., 2002). The USA National Polyp Study (NPS) reports a 70-90% lower than expected incidence of colorectal cancer in patients undergoing colonoscopic surveillance (Winawer et al., 1993). Several studies have shown that this risk is related to the characteristics of previously removed adenomas (Atkin and

Saunders, 2002). Current UK adenoma surveillance guidelines are based on the size and number of polyps removed (**Figure 5**) (Atkin and Saunders, 2002).

e) NHS Bowel Cancer Screening Programme (NHS BCSP)

Regular bowel cancer screening has been shown to reduce the mortality risk from bowel cancer by 16% (Hewitson et al., 2007). Colorectal cancer lends itself well to population screening because it is common, has a well-recognised premalignant lesion and treatment of the premalignant lesion reduces the risk of cancer (West et al., 2008). The NHS Bowel Cancer Screening Programme (NHS BCSP) commenced in April 2006 and invites men and women aged 60–74 to participate via submission of a Faecal Occult Blood test (FOB) every 2 years; those with a positive result will be offered colonoscopy as the next investigation of choice (West et al., 2008). If adenomatous polyps are detected, they will be removed and the patient will undergo colonic surveillance.

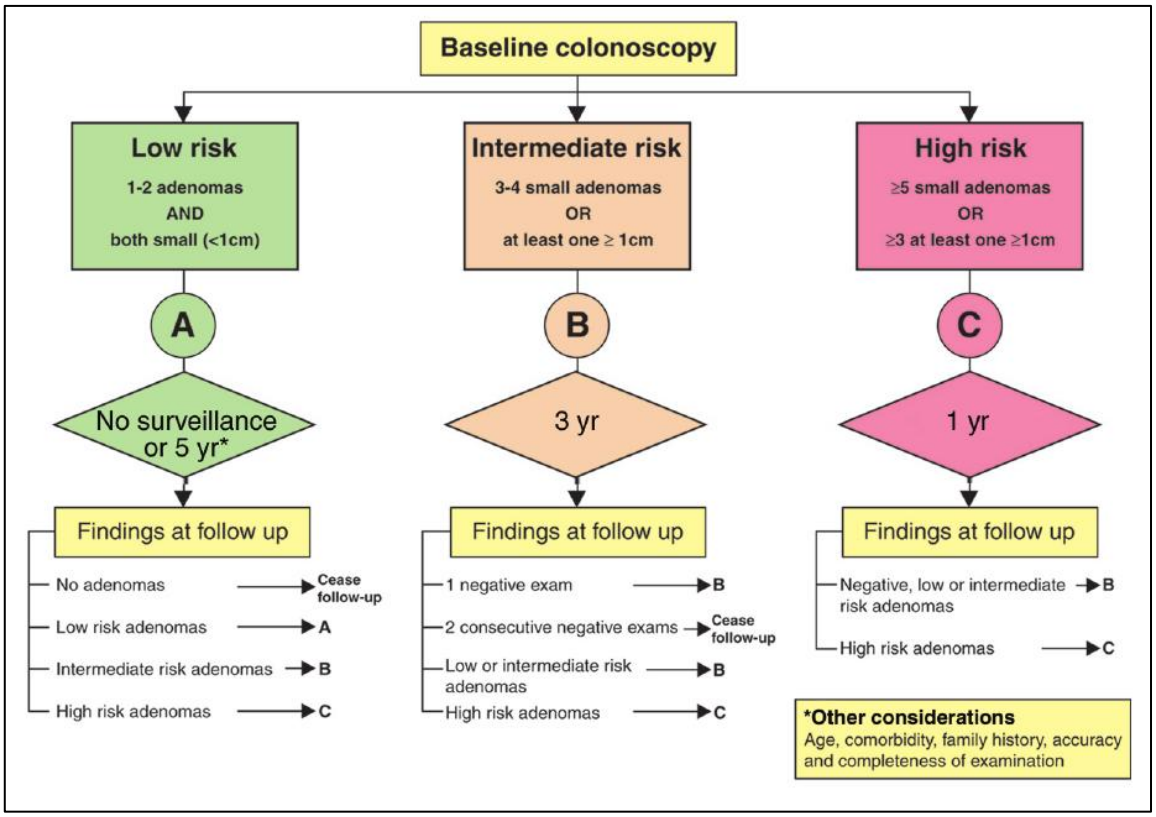


Figure 5: UK polyp surveillance following adenoma removal (Atkin and Saunders, 2002)

The NHS BCSP has led to a rise in the adenomatous polyp detection rate and a subsequent demand to train more endoscopists to manage this increasing workload. Screening centres each serve a population of between 500,000–2 million people. Assuming a 60% uptake and 2% positivity (based on the pilot study), this would necessitate an estimated 300 colonoscopies per year, equating to one or two extra endoscopy lists per hospital, per week (West et al., 2008, Endoscopy., 2000). One way of addressing this increased demand is to train more nurse endoscopists. The nurse practitioner flexible sigmoidscopy programme was first introduced in 1996 (Duthie et al., 1998). Subsequent studies have shown that there is no difference in effectiveness or patient satisfaction for flexible sigmoidscopy performed by a registered nurse, general surgeon or gastroenterologist (Schoenfield et al., 1999). In 2007, the British Healthcare Commission reported that 85% of acute hospital trusts employed nurse endoscopists leading to reduce waiting times for investigations and a reduction in the outpatient workload (Health. 2007). More recently, feasibility studies have been conducted which demonstrate that nurse endoscopists perform colonoscopy according to internationally recognised standards with high patient satisfaction (Van Putten et al., 2012).

1.2 Colonoscopic polypectomy

Colonoscopic polypectomy was established by Shinya and Wolfe during the 1970's and is now a widely accepted and practised technique (Sivak, 2004). There are several different methods of polypectomy that can be employed depending on the nature and position of the polyp being removed.

a) Hot and Cold Biopsy

Diminutive polyps (≥3mm) may be removed by cold biopsy (without diathermy) irrespective of morphology from anywhere in the colon (Riley, 2008). This is the simplest method for

polypectomy allowing cold forceps to grasp small polyps that may otherwise be too small to snare (Fyock and Draganov, 2010) (**Figure 6**). Advantages of cold forceps polypectomy include avoiding risks associated with electrosurgery and an almost negligible risk of colonic perforation (Rex, 2010). Disadvantages include firstly, that it is likely to leave residual tissue unless the endoscopist is particularly vigilant. Secondly, it is inefficient when more than 2 or 3 biopsies are required and thirdly, the field may become obscured with blood with subsequent biopsies necessitating flushing (Riley, 2008). Despite this, results from a recent multicentre prospective study have shown the high safety of this approach with low rates of post procedural bleeding (Repici A et al., 2012).

Hot biopsy uses electrosurgery to destroy the residual polyp tissue left behind (Gilbert et al., 1992). During this technique only the tip of the polyp is grabbed in the forceps. The polyp is pulled into the colonic lumen to create a tent-like effect and electrocautery is applied to destroy the polyp base while preserving the polyp tissue inside the forceps as a histological specimen (Williams, 1991). Due to the risks of transmural thermal injury, it is best avoided in the right colon where the wall is thin (Riley, 2008). Metz *et al* have shown that in a porcine model hot biopsy results in a significantly greater depth of tissue injury, with a high proportion of transmural necrosis (Metz et al., 2013). The authors conclude that hot biopsy is imprecise, potentially ineffective and hazardous and for these reasons, it is now becoming less commonly used (Metz et al., 2013).

b) Snare polypectomy

Snare polypectomy is the preferred method for removal of polyp's ≥1 cm in size (**Figure 7**) (Singh et al., 2004). A snare is opened over the polyp and then closed entrapping the tissue for resection (Fyock and Draganov, 2010). Once the polyp is captured, the snare's plastic sheath is advanced, moving the polyp away to avoid electrosurgical damage to the

colonoscope (Fyock and Draganov, 2010). The snare is placed approximately half way up the stalk, so that after cutting, a remnant is left which can be grabbed or clipped if haemorrhage occurs (Fyock and Draganov, 2010). Before diathermy is applied, the polyp is pulled away from its base into the lumen tenting the colon wall to avoid burning the adjacent deep colonic layers (Tolliver and Rex, 2008).

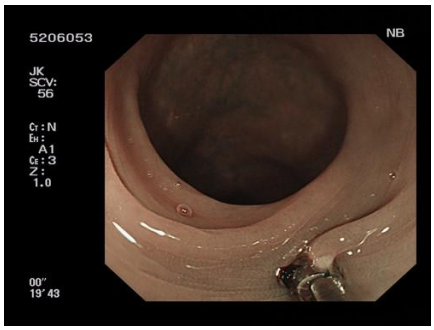


Figure 6: Endoscopic view of cold biopsy (Source, 2013)

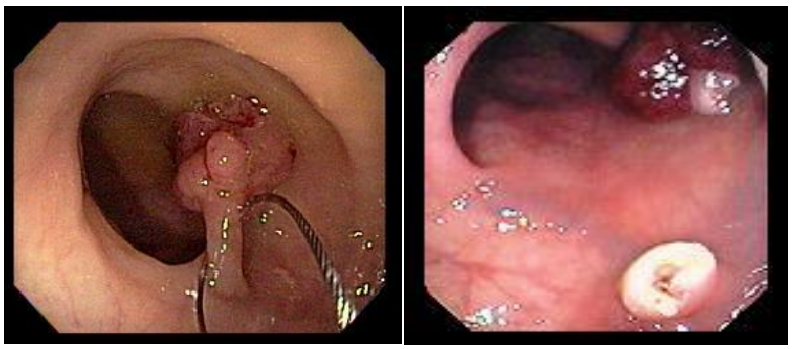


Figure 7: Endoscopic view of snare polypectomy (left = pre and right = post) (Canada, 2013)

c) Endoscopic Mucosal Resection (EMR) / Endoscopic Submucosal Dissection

Endoscopic Mucosal Resection (EMR) (**Figure 8**) can be performed on sessile polyps ≥2 cm (Fyock and Draganov, 2010). EMR uses submucosal injection to create a cushion for the polyp and then hot snaring to remove the polyp either *en bloc* or piecemeal (Fyock and Draganov, 2010). Raising the polyp can also identify lesions invading or tethered to the deep submucosa or muscle layer (the non-lifting sign) which are unlikely to be suitable for

endoscopic removal (Fyock and Draganov, 2010). Endoscopic Submucosal Dissection (ESD) aims to remove all dysplastic tissue en-bloc as one piece rather than the piecemeal technique (Kantsevoy et al., 2008). The ideal submucosal injection solution should provide a high cushion to aid safe resection whilst also preserving tissue for histopathological assessment (Uraoka, T. et al., 2008). Many endoscopists use saline with or without adrenaline but a wide variety of solutions are available and may result in longer lasting cushions (Riley S et al. 2008). These includes high viscosity solutions such as glycerol, dextrose water, hyaluronic acid, carboxymethylcellulose and chitosan hydrogel (Uraoka, T. et al., 2008). An injection proximal to the lesion may help tilt it towards the endoscope making it easier to snare (Riley S et al. 2008). Before performing the EMR, the endoscopist should clearly identify the margins of the lesion in order to avoid incomplete resection. The application of indigo carmine (**Figure 9**) may help and some recommend marking the periphery of the lesion with electrocautery (Riley S et al. 2008).

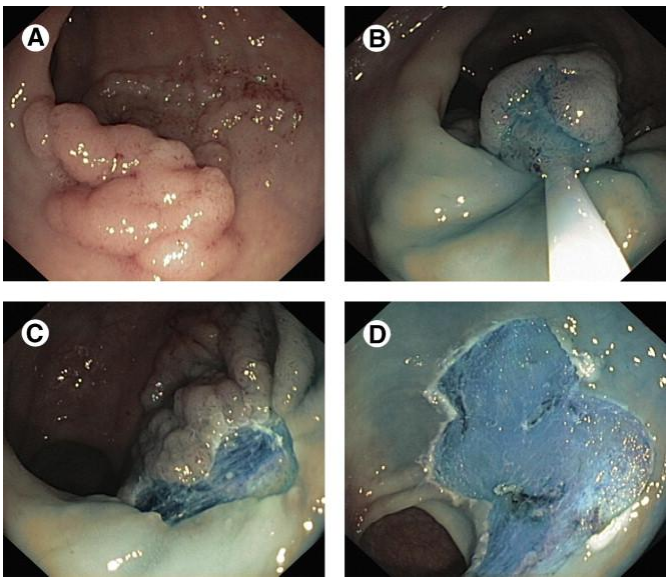


Figure 8: Endoscopic Mucosal Resection (EMR) A = Paris 0-11a granular laterally spreading tumour, B = first snare excision, C = exposed submucosa, D = complete snare excision (Bourke, 2011)

d) Complications following colonoscopic polypectomy

Colonoscopic polypectomy significantly reduces the risk of colon cancer development however; the procedure is not without risk. Most complications are related either to post-polypectomy haemorrhage or perforation. The risk of haemorrhage ranges from 0.3% to 6% but can be as high as 24% in large polyps (Rosen et al., 1993). Post polypectomy haemorrhage (**Figure 10**) is usually divided into immediate (≤12 hours post-procedure) and delayed (>12 hours post-procedure) (Rosen et al., 1993). A recent multicentre study demonstrated that difficult colonic polypectomy is unpredictable with post procedural haemorrhage being independent of polyp type or size (Voiosu T et al., 2013). Immediate bleeding can be managed by the application of pressure with a 1:10,000 dilution of adrenaline, thermal therapy, endoscopic haemoclips or band ligators (Hong et al., 2012). Delayed bleeding is usually self limiting and resolves with supportive care in 70% of cases but persistent bleeding may require similar endoscopic interventions as immediate bleeds (Farrell J.J et al., 2005).

Perforation can result from mechanical stress, barotrauma, electrocautery, and the depth of the polyp resection itself (Fyock and Draganov, 2010). The risk of perforation with all colonoscopies has been estimated somewhere around 1 per 1000 to 2000 (Fyock and Draganov, 2010). Perforation following polypectomy can be managed by both operative and non operative strategies (Lohsiriwat et al., 2010). Conservative treatment is reserved for patients with small perforations, without signs of peritonitis (Lohsiriwat et al., 2010). This involves intravenous fluid, antibiotics and absolute bowel rest. The overall success rate is highly variable (33%-73%) (Lohsiriwat et al., 2010). A second option is closure of the perforation using endoscopic clips. In general this is reserved for perforations less than 1cm in size (Lohsiriwat et al., 2010). These patients require intravenous antibiotics and clear

liquid diet until bowel movements return and evidence of peritonitis disappear (Lohsiriwat et al., 2010). Finally, surgical management is required for any patient with either diffuse peritonitis, clinical deterioration following conservative treatment, or those with cancerous colonic pathology (Lohsiriwat et al., 2010). Small perforations with minimal faecal contamination can be over sewn but bowel resection is required for large perforations or cancerous lesions (Lohsiriwat et al., 2010). Approximately 5% of perforations result in patient death (Fyock and Draganov, 2010). It is therefore extremely important that polypectomy training is conducted in a well regulated and structured manner.

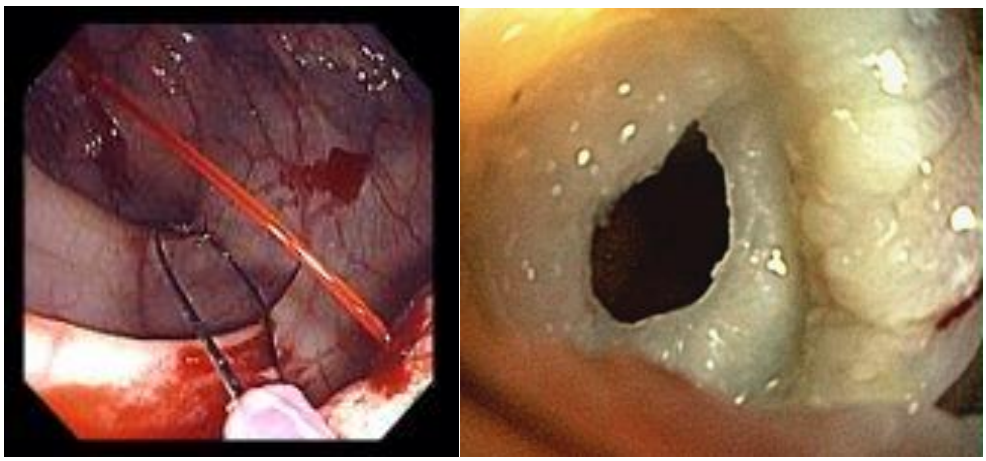


Figure 9: Complications following colonoscopic polypectomy. Left = acute haemorrhage following polypectomy, Right = post polypectomy perforation (Lin et al., 2009).

1.3 Training and assessment in colonic polypectomy

a) Colonoscopy certification and current assessment procedure

In view of the increasing adenoma detection rate due in part to the NHS BCSP and the multiple techniques available for polypectomy, appropriate training and assessment is of paramount importance. Endoscopic training in the United Kingdom (UK) is governed by the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) and administered through the

JAG Endoscopy Training System (JETS) (JAG, 2012). JAG was established in 1994 to set acceptable standards for endoscopy units and to assure quality in endoscopic training (Physicians, 2012). Trainee endoscopists must complete several "eligibility criteria" for provisional and full certification in colonoscopy (**Tables 2 and 3**).

The need for improved training in colonoscopy practice was demonstrated by a prospective study of colonoscopy practice in 2004 (Bowles et al,. 2004). A total of 9223 colonoscopies in 68 UK endoscopy units were evaluated with an adjusted caecal intubation of 56.9%. Only 17.0% of colonoscopists had received supervised training for their first 100 colonoscopies and just 39.3% had attended a training course (Bowles et al,. 2004).. This has improved significantly since the introduction of JAG accreditation. The 2013 UK national colonoscopy audit now reports an adjusted caecal intubation rate of 95.8% for more than 20,000 colonoscopies (Gavin D.R et al., 2013).

Eligibility criteria	Requirements	Notes
Caecal Intubation rate	>90%	-
Formative DOPS[1] scores	>90% "3"s and "4"s	Minimum 10 required
Formative DOPyS[1] scores for polypectomy level 1	>90% "3"s and "4"s	Last 4 DOPyS scores, or last 3 months, whichever greater
Unassisted physically	>90%	Trainer does not take the scope for >90% of procedure
Basic skills lower GI course	Attended	-

Table 2: Criteria for provisional certification in colonoscopy (JAG, 2011). DOPS = Directly Observed Procedural Skills, DOPyS = Direct Observation of Polypectomy Skills, level 1 = polyps <1cm in size, level 2 polyps >1cm in size

[1] DOPS / DOPyS are qualitative assessments with multiple parameters marked on scale of 1-4, see section 1.3*b*

Eligibility criteria	Requirements	Notes
Caecal Intubation rate	>90%	-
Polyp detection &removal	>10%	-
Serious complications	<0.5%	Death, perforation, significant bleeding requiring >2unit transfusion, post-procedure hospital stay >24hrs or hospital admission due to a procedural complication following discharge
Sedation rates	Mean below recommended	<70yrs < 5mgs midazolam and <50mgs pethidine/100ug fentanyl >70yrs < 2.5mgs midazolam and <25mgs pethidine/50ug fentanyl
Formative DOPyS scores	>90% “3”s and “4”s	For polypectomy level 2

Table 3: Criteria for full certification (JAG, 2011)

b) *Subjective assessment in colonoscopic polypectomy*

The Direct Observation of Polypectomy Skills (DOPyS) (**Figure 11**) is a qualitative assessment process used for JAG accreditation (Gupta et al., 2011, Gupta et al., 2012). There is evidence to suggest that this is a valid and reliable form of assessment (Gupta et al., 2012). Research has shown the DOPyS to accurately differentiate between polypectomies performed by endoscopists of different levels of experience (Gupta et al., 2012). It is marked on a scale of 1 to 4 where, *4 denotes a highly skilled performance, 3 a competent performance, 2 where some standards are not yet met, aspects to be improved, some errors uncorrected and 1 where accepted standards are not yet met with frequent uncorrected errors.* Scores are awarded against a list of 34 parameters which are divided into generic, stalked polyps, sessile polyps and post polypectomy categories. Polyps are defined as level 1 (<1cm) and level 2 (>1cm). For provisional certification, trainees must demonstrate competency in level 1 polypectomy and for full certification in level 2 polypectomy.

c) *Objective assessment in colonoscopic polypectomy*

One limitation of the DOPyS is that there is an element of subjective interpretation in the assessment process. Objective parameters of polypectomy performance are not currently available. Obstein *et al* highlights the potential use of kinematics in order to evaluate technical skill during simulated colonoscopy (Obstein et al., 2011). The authors attached electromagnetic sensors onto a colonoscope to measure path length, tip-angulation, absolute roll and scope curvature to quantify performance (Obstein et al., 2011). Differences in these parameters were found according to the expertise of each participant to provide a mean kinetic score for overall performance (Obstein et al., 2011). Clark *et al* has used similar techniques to develop a quantitative scale of endoscopic torque control during Natural Orifice

Transluminal Endoscopic Surgery (NOTES) using motion tracking of wrist movements (Clark et al., 2013). Similar studies have yet to be performed for colonic polypectomy.

DOPyS: Polypectomy Assessment Score Sheet

Date ...../...../..... Assessor..... Colonoscopist: ..... Case ID:..... Polyp Number.....

Polyp site: C / AC / HF / TC / SF / DC / SC / R

Scale:
- **4** - **Highly skilled performance**
- **3** - **Competent & safe throughout procedure, no uncorrected errors**
- **2** - **Some standards not yet met, aspects to be improved, some errors uncorrected**
- **1** - **Accepted standards not yet met, frequent errors uncorrected**
- **N/A** - **Not applicable/Not assessable**

The underlined parameters can only be assessed during 'live' polypectomy

Generic	Score	Comments
Optimising view of / access to the polyp:		
1. Attempts to achieve optimal polyp position		
2. Optimises view by aspiration/insufflation/wash		
3. Determines full extent of lesion (+/- use of adjunctive techniques e.g. bubble breaker, NBI, dye spray etc) if appropriate		
4. Uses appropriate polypectomy technique (e.g. taking into account site in colon)		
5. *Adjusts/stabilises scope position*		
6. *Checks all polypectomy equipment (forceps,snare,clips,loops) available*		
7. *Checks (or asks assistant to) snare closure prior to introduction into the scope*		
8. *Clear instructions to and utilisation of endoscopy staff*		
9. *Checks diathermy settings are appropriate*		
10. *Photo-documents pre and post polypectomy*		
Stalked polyps: Generic, then		
11. Applies prophylactic haemostatic measures if deemed appropriate		
12. Selects appropriate snare size		
13. Directs snare accurately over polyp head		
14. Correctly selects en-bloc or piecemeal removal depending on size		
15. Advances snare sheath towards stalk as snare closed		
16. Places snare at appropriate position on the stalk		
17. Mobilises polyp to ensure appropriate amount of tissue is trapped within snare		
18. Applies appropriate degree of diathermy		
Small sessile lesions / Endoscopic mucosal resection: Generic, then		
19. Adequate sub mucosal injection using appropriate injection technique, maintaining views		
20. Only proceeds if the lesion lifts adequately		
21. Selects appropriate snare size		
22. Directs snare accurately over the lesion		
23. Correctly selects en-bloc or piecemeal removal depending on size		
24. Appropriate positioning of snare over lesion as snare closed		
25. Ensures appropriate amount of tissue is trapped within snare		
26. Tents lesion gently away from the mucosa		
27. Uses cold snare technique or applies appropriate diathermy, as applicable		
28. Ensures adequate haemostasis prior to further resection		
Post polypectomy		
29. Examines remnant stalk/polyp base		
30. Identifies and appropriately treats residual polyp		
31. Identifies bleeding and performs adequate endoscopic hemostasis if appropriate		
32. Retrieves, or attempts retrieval of polyp		
33. *Checks for retrieval of polyp*		
34. Places tattoo competently, where appropriate		

Polyp Size	.....mm		
Overall Competency at Polypectomy:	**4**	**3**	**2**	**1**

Figure 10: Direct Observation of Polypectomy Skills (DOPyS) (Gupta et al., 2011)

d) Self-assessment

Another area of assessment which has not been fully explored in endoscopy is self-assessment A meta-analysis of 44 self-assessment studies in medical education reported a moderate correlation between self and expert assessments (Falchikov NB, 1989). A similar review by Gordon *et al* of 18 papers demonstrated comparable findings (Gordon, 1991). There are some reports in the literature that the ability to accurately self-assess improves with experience because the participant can recognise an expert performance and use this as a benchmark to assess their own skills (Moorthy et al., 2006, Ward et al., 2003). Advanced and expert colonoscopists may be familiar with assessing novice and intermediate performances but may be less able to repeat this process for themselves.

1.4 The use of simulation in colonoscopy training

The majority of JAG training courses utilise simulation as an adjunct to "real-life" training. Over the past decade, there has been an explosion in the development and utilisation of training models and simulators for endoscopic training (Sedlack et al., 2004). Simulation is particularly attractive in the field of endoscopy because it avoids the use of patients for skills practice and ensures that trainees have had some exposure before treating humans (Issenberg et al., 1999). It also allows the participant to gain the skills needed to progress along their learning curve in a safe environment (Reynolds and Kong, 2010, Stather et al., 2011). There are several different types of colonoscopy simulation available for training (**Table 4**). However, there are no simulators which focus solely on colonoscopic polypectomy.

Simulator	Examples	Description
1. Mechanical	- Adam Rouilly simulator / Koken model I-B	Silicone rubber colonic model for basic colonoscopy procedure. Lumen of simulated bowel shaped in order to replicate different areas of the colon (Koken, 2012)
	- Chamberlain Group LLC	Plastic colon, mounted in rigid foam. Has the ability to be fitted with replaceable colonic stricture and polyps. This group also make a straight colon section to train polypectomy and stenting (Group, 2011)
2. Composite animal	- Endo X trainer ™ (Medical Innovations)	Portable plastic tray system, can lay animal tissues within the tray and can perform a variety of procedures without or without simulated bleeding (Desilets et al., 2011, Sedlack et al., 2007b)
	- Colonoscopy suitcase (WIMAT)	Portable colonoscopy trainer which can simulate removal of sessile and pedunculated polyps with/without the capacity for bleeding and diathermy
3. Virtual reality	- Accutouch HT (CAE Previously Immersion Medical)	Trolley-mounted, computerised device with a flat-screen display on a movable arm. A model endoscope is provided with the system. Several modules are available range of endoscopic procedures and pathology. It simulates patient vital signs and responses to administration of sedation and to pain (Desilets et al., 2011)
	- GI mentor II (Simbionix)	Contains library with over 120 tasks. Can measure end points including procedure time, visualisation of bowel, mechanical pressures on bowel. Has modules for degrees of difficulties and dealing with pathology (Simbionix, 2012)
	- KAIST-Ewha	Manages training scenarios with varying degrees of difficulty, measures multiple parameters including; time taken, exertion force, tip motion, number of red outs (Woo et al., 2008b)
	- Endo TS1 (2nd Gen) (Olympus)	For training and assessment of colonoscopy skills. Simulates multiple matrices including; shaft looping, tip contact, variable shaft stiffness (Haycock et al., 2009a)
4. Live animal models	- Porcine model	A realistic platform with haptic feedback similar to human tissue. These are expensive with ethical concerns and can demonstrate anatomical variation

Table 4: Summary of colonoscopy simulators available for training

a) Mechanical simulators / box trainers

The Erlangen Plastic Mannequin for upper endoscopic management is the earliest endoscopic mechanical simulator to be reported (Classen and Ruppin, 1974). There are now several manufacturers of this type of simulator (Group., 2012, Koken, 2012, Rouilly, 2012, Erlangen, 2012, Limbs and Things, 2012) (**Figure 12**). Although some advances have been made, mechanical models lack realism because of poor simulation of tissue properties (Desilets et al., 2011). There is also little variety, which serves to limit their usefulness (Desilets et al., 2011). These simulators have largely been replaced by more realistic and adaptable alternatives but can still be used for the novice during the early stages of learning (Desilets et al., 2011).

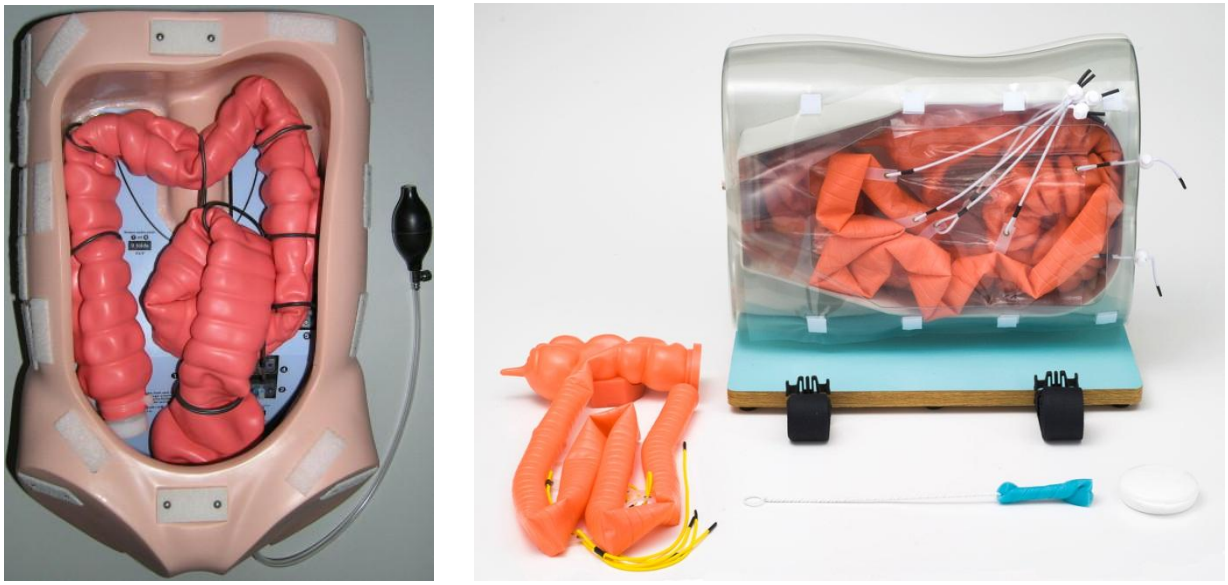


Figure 11: Left = Kyoto Kagaku Colonoscope Training Model (Kyoto Kagaku, Japan) (Plooy et al., 2012), Right = Colonoscopy & Enteroscopy Training Simulator, (Buyamag, USA)

b) Live Animal Simulators

Porcine models have been adopted as the live animal model of choice for endoscopic simulation (Nelson et al.) The haptic feedback is similar to human tissue, although the thickness and orientation of various organs can be different (Desilets et al., 2011) . In the United Kingdom (UK), the Cruelty to Animals Act (1876) prevents their use (Cruelty to Animals Act, 2012). This type of simulation requires speciality units with considerable financial investment including pre-operative preparation, sedation, induction and anaesthetic monitoring (Desilets et al., 2011). Rigid adherence to the requirements of the Animal Welfare and Ethics Committee also is essential (Desilets et al., 2011).

c) Virtual Reality Simulators

Virtual reality (VR) computer simulators are becoming a popular way of providing trainees with an opportunity to practice endoscopy (Walsh et al., 2012). These utilise computer based modules with varying degrees of difficulty. This enables the participant to perform multiple simulations and record results to monitor their progress. There are currently several VR endoscopic trainers available (**Figure 13**) (Simbionix, 2012, Olympus, 2012, Kim et al., 2007, Long and Kalloo, 2006). The major limitation of VR training is the financial expense with the average cost of a VR trainer being in excess of £100,000 (Simbionix, 2012). Another criticism is that most VR are too simplistic to be beneficial for anyone other than the novice trainee and that the outcomes of assessment may be irrelevant as markers of expertise (Sedlack et al., 2004).

d) Ex-vivo animal simulators

Ex-vivo simulators are constructed from a combination of plastic parts and explanted animal organs. These have overcome some of the limitations of live animal models (Desilets et al.,

2011). The Erlangen Active Simulator for Interventional Endoscopy (EASIE) was one of the first models of this type to be developed for endoscopic training (Hochberger et al., 1997). This device consists of a plastic head and torso mounted on a tilting device (Desilets et al., 2011). Porcine upper gastro intestinal organs are inserted into the simulated abdomen and an arterial perfusion system feeds in synthetic coloured fluid via an electric pump to simulate arterial bleeding (Desilets et al., 2011). This model has been further developed in to make it lighter and more portable, this is known as the compactEASIE (Hochberger et al., 2005). A bovine model for diagnostic colonoscopy, the Endo X Trainer (Medical Innovations International, Rochester, Minn) (**Figure 14**) has also recently been developed (Sedlack et al., 2007a). This is designed to allow the trainee to practice colonoscopic navigation.

There are several benefits of ex-vivo simulation. Models are adaptable and can be used to simulate a range of endoscopic procedures (Desilets et al., 2011). They can provide a more realistic feel compared with purely mechanical models. Ex-vivo models also enable the practice of therapeutic endoscopy in a controlled setting. The low cost makes it a financially viable option for training large numbers of participants with varying levels of experience.

e) The WIMAT colonoscopy suitcase[†]

The WIMAT colonoscopy suitcase is an ex-vivo porcine simulator designed to teach the principles of the colonic polypectomy. This model has been designed at the Welsh Institute for Minimal Access Therapy (WIMAT) (**Figure 13**). This novel simulator has the potential to simulate a range of polypectomy procedures.

[†] Details of the construction of the WIMAT colonoscopy suitcase are outlined in Chapter 3

Figure 12: Examples of VR colonoscopy simulators. Left = The Simbionix GI mentor II (Simbionix, USA) (Simbionix, USA), Right = LSRO Colonoscopy Simulator (LSRO, USA)

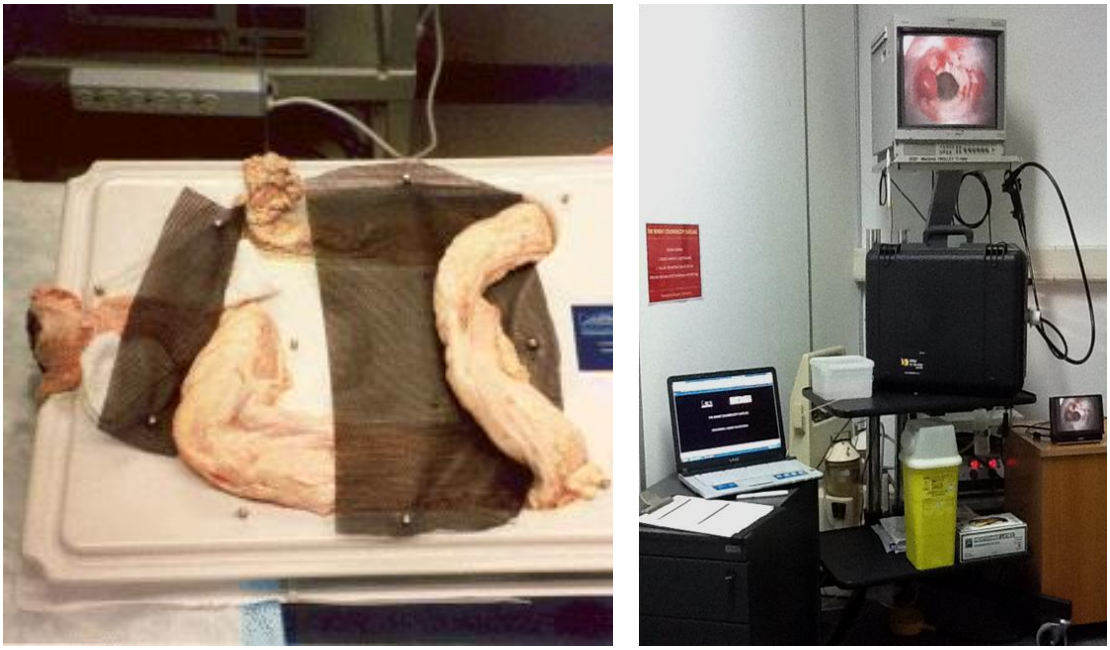


Figure 13: Examples of Ex-vivo colonoscopy simulators. Left = Ex-vivo bovine colonoscopy simulation model (Endo X, USA) (Sedlack et al., 2007b), Right = The WIMAT colonoscopy suitcase

f) The requirements of a training simulator

"A surgeon trained on a simulator is twice as fast and twice as accurate as one who has not been. It reduces errors, making surgery much safer. The NHS must be able to provide it to make a difference to patients" Sir Liam Donaldson, 2009

There is good evidence to suggest that this is an accurate reflection of simulator training (Sedlack et al 2011). Clinicians that utilise simulators as an adjunct to training tend to perform subsequent procedures in less time and with fewer errors (Sedlack et al 2011). Although the long term follow up data does not exist, inferences can be made that simulator training may improve patient safety. This cannot however, be assumed for all types of simulation and therefore, with the development of any new simulator, validation of its effectiveness is a critical process (Sedlack, 2011). Validation is the extent to which an instrument measures what it was designed to measure (Desilets et al., 2011). Before a simulator can be used for either training or assessment, evidence of its validity is required (Plooy et al., 2012). There are several different facets of simulator validity testing (**Table 5**) and the more aspects of validity proved, the stronger the argument (Sedlack, 2011).

Validation	Definition	Method
Face	The rational expectation that an association between two things exists	*Questionnaire to non-expert users*
Content	The extent to which a measure reflects the trait or domain it purports to measure. For training tool: need data to prove haptic and visual realism. For assessment tool: review by experts of the skill domains being tested	*Questionnaire to expert users*
Construct / Contrast	An agreement between a theoretical concept and a specific tool or procedure (Experienced practitioners should score higher on its assessment parameters than juniors)	*Measuring relevant parameters for defined groups of variable expertise*
Criterion *(Predictive / Concurrent)*	*Predictive validity*: The ability of a tool to predict future performance	*Correlation of test score with future performance*
	Concurrent validity: The correlation between assessment tool and the "gold standard"	*Comparison with patient based data*

Table 5: Types of validity testing for simulation models (adapted from: Sedlack, 2011)

g) Summary

- The demand for colonoscopic polypectomy has increased following the BCSP
- Colonoscopic polypectomy has a long learning curve with significant risk of complications to the patient
- Polypectomy training therefore needs to be carried out in a safe and controlled manner
- Simulation has the potential to address these issues, but at present, current endoscopic simulators lack the realism and appropriate level of difficulty needed
- Ex-vivo simulation may provide a solution and therefore, a new simulator called the WIMAT colonoscopy suitcase has been developed

1.5 Aims and Hypotheses

a) Aims

1. To develop an existing ex-vivo porcine simulator for colonoscopic polypectomy training
2. To test the content, construct and criterion validity of the simulator
3. To evaluate the feasibility of measuring skills transfer (from the simulator to the clinical environment) using the simulator
4. To evaluate if self-assessment is an accurate when using the simulator
5. To establish a quantitative scoring system for colonic polypectomy using the simulator

b) Hypotheses

- The simulator is valid tool for colonic polypectomy skills training
- Measuring skills transfer to the clinical setting following simulator training is feasible
- Trainees can use the simulator to accurately self-assess their performance
- The simulator can be used to establish a quantitative scoring system

Chapter 2: Review of published literature

A systematic review of validity testing in colonoscopic simulation

2.1 Introduction

Establishing the validity of a training simulator is a critical process in its development and subsequent adoption. Using poorly validated models may lead to the inaccurate reporting of trainee performance levels (Van Nortwick et al., 2010). There are numerous papers in the literature that address endoscopic simulator validation but no review evidence which focuses solely on colonoscopy and therapeutic colonoscopy. Between August 2011 and January 2012 a systematic review was conducted in order to identify the evidence for validity in this field. The aim of this review was to evaluate the strength of this research and highlight areas of colonoscopy simulator validation which have yet to be explored.

2.2 Methodology

The review was conducted according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines (PRISMA, 2011). Scientific databases were searched which included; Embase classic and Embase (1947-2011), Medline (1947-2011), PubMed (1966-May 2010), *meta*Register of Controlled Trials and Education Resources Information Centre (ERIC). Three different domains of exploded Medical Subject Headings (MeSH) terms were used (**Appendix 1**). The first domain contained multiple terms for endoscopy, the second for simulation and the third for validation. All studies validating colonoscopy simulators for assessment or training were included in the review and to ensure reliability, a second investigator repeated the search. Any differences in opinion were referred to a third party for final analysis.

a) Inclusion and exclusion criteria

Included studies needed to contain sufficient details of the simulation model used and type of validity measured. Reviews, congress abstracts and studies that validated tools of assessment and not the simulator itself were excluded. Studies detailing endoscopic methods other than colonoscopy were also excluded.

b) Outcome measures

For each paper, the type of simulator used was recorded along with the task being assessed, the endpoints of the study and the type of validity measured. The common endpoints between papers were compared when statistically significant results were reported. The principle summary statistic was the difference in means. The heterogeneity of the methodology prevented meta-analysis.

2.3 Results

a) Search strategy and findings

The primary search identified 1,141 studies. After duplicates were removed, 739 titles and abstracts were screened for relevance. From this, 678 records were excluded, leaving 61 full text articles for review. Papers deemed not eligible at this stage were usually referring to non lower GI simulators. After review of the full text, 53 records were excluded and 5 cross-referenced, leaving a total of 13 papers for inclusion (**Figure 15**). All included papers reported the evidence for the validity of colonoscopy simulation (**Table 6**). Twelve (92.3%) of these articles focussed on VR simulator validation and 1 (7.7%) on ex-vivo simulator validation. There were no papers which validated a colonoscopic polypectomy simulator.

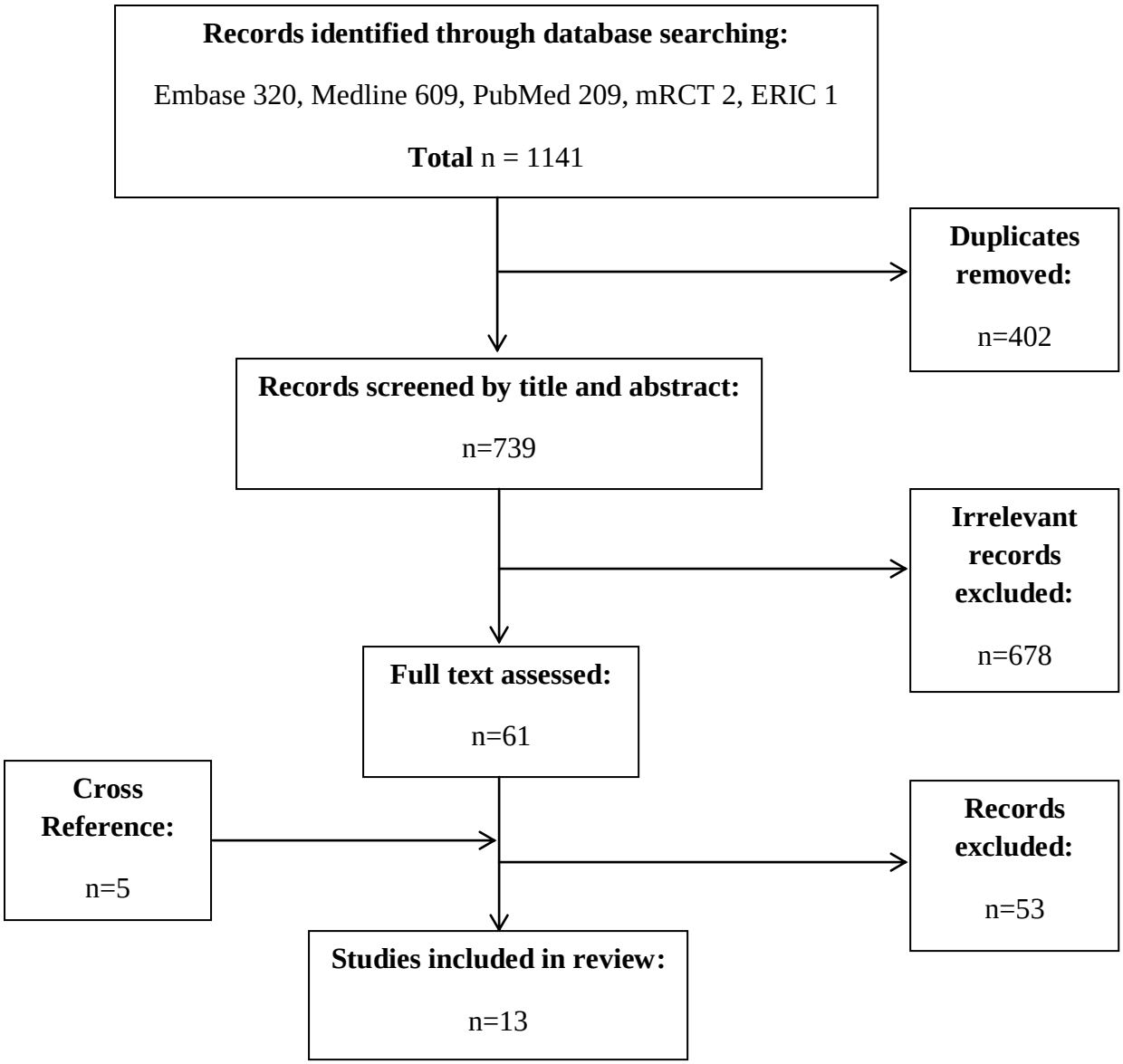


Figure 14: Summary of the search strategy and findings

b) The type of simulators included in the review

The 12 (92.3%) VR simulator validation studies included; 5 (41.6%) which evaluated the Accutouch HT immersion (Immersion Medical, Germany), 4 (33.3%) which evaluated the Simbionix GI mentor II (Simbionix, USA), 2 (16.6%) which evaluated the Olympus 2nd Generation (Olympus, Japan) and 1 (8.3%) which evaluated the KAIST-Ewha (KAIST,

Korea). One novel study looked at a composite model using bovine intestine and the Endo-X colonoscopy platform (Medical Innovations International, USA).

Study	Simulator tested	Simulator
(Datta et al., 2002)	Accutouch HT immersion (Immersion Medical, Germany)	VR
(MacDonald et al., 2003)	Accutouch HT immersion (Immersion Medical, Germany)	VR
(Mahmood and Darzi, 2003)	Accutouch HT immersion (Immersion Medical, Germany)	VR
(Sedlack and Kolars, 2003)	Accutouch HT immersion (Immersion Medical, Germany)	VR
(Moorthy et al., 2004)	Accutouch HT immersion (Immersion Medical, Germany)	VR
(Felsher et al., 2005)	Simbionix GI mentor II (Simbionix, USA)	VR
(Grantcharov et al., 2005)	Simbionix GI mentor II (Simbionix, USA)	VR
(Koch et al., 2008a)	Simbionix GI mentor II (Simbionix, USA)	VR
(Fayez et al., 2010)	Simbionix GI mentor II (Simbionix, USA)	VR
(Koch et al., 2008b)	Olympus 2nd Generation (Olympus, Japan)	VR
(Haycock et al., 2009a	Olympus 2nd Generation (Olympus, Japan)	VR
(Woo et al., 2008b)	KAIST-Ewha (KAIST, Korea)	VR
(Sedlack et al., 2007b)	Novel bovine model (Endo-X, USA)	Ex-vivo

Table 6: Summary of studies meeting inclusion criteria for the review

c) Evidence for the face / content validity of colonoscopy simulators

Five studies included in the review provided evidence for face or content validity. This type of validity was measured by the use of Likert-scale realism scores. Face validity was confirmed in 2 studies for the Olympus Endo 2nd Generation (Olympus, Japan) (Haycock et al., 2009a, Koch et al., 2008b) and in two studies for the Accutouch HT immersion simulator (Immersion Medical, Germany) (Mahmood and Darzi, 2003, Sedlack and Kolars, 2003). One study reported the face validity for the KAIST-Ewha (KAIST, Korea) (Woo et al., 2008a).

Content Validity (obtained from experts completing realism surveys, (**Table 7**) was reported in 1 study each for the Simbionix GI mentor II (Simbionix, USA) and the Accutouch HT immersion (Immersion Medical, Germany) (Koch et al., 2008a, Sedlack and Kolars, 2003). Content validity was also reported in 2 studies for the Olympus 2nd Generation (Olympus, Japan) (Haycock et al., 2009a, Koch et al., 2008b). One study reported the content validity of the Endo X ex-vivo model (Endo-X, USA) (Sedlack et al., 2007b). There were no reports of criterion validity for any VR simulators.

d) Evidence for the construct validity of colonoscopy simulators

Construct validity was the most widely reported form of validation for colonoscopy simulation. In all studies examining construct validity, participants were assigned to groups according to the previous number of colonoscopies they had performed. Each study had different criteria for this (**Table 8**). Eleven (84.6%) studies focused on VR and 1 (7.7%) on ex-vivo simulation. Participants were asked to complete a variety of modules/cases assessing navigation, diagnostics or therapeutics. The nature of these tasks also varied. Several common endpoints were comparable between studies that validated the same simulator (**Table 9**)

Simulator	Study	Survey scale	Measured Parameters	Likert Value
Accutouch HT (Immersion Medical, Germany)	(Sedlack and Kolars, 2003)	10-Point	Realism of controls	7.9
			Visual graphics	7.0
			Force / Feel	6.5
			Insufflation and suction	5.7
			Scope controls	8.0
			Loop management	6.6
Olympus 2nd Generation (Olympus, Japan)	(Koch et al., 2008b)	10-Point	Difficulty	7.2
			Practical set-up	6.9
			Endoscopic handling	7.6
			Endoscope movement	7.1
			Tactile feedback	7.0
			Insertion	7.1
	(Haycock et al., 2009a)	10-Point	Appearance	5.8
			Movement	6.4
			Force feedback	6.6
			Looping	6.6
			Loop resolution	6.8
Simbionix GI mentor II (Simbionix, USA)	(Koch et al., 2008a)	4-Point	Overall realism	3.0
			Anatomical Representation	2.6
			Simulator set-up	3.1
			Endoscopic control	3.2
			Haptic feedback	2.6
Bovine model (Endo-X, USA)	(Sedlack et al., 2007b)	7-Point	Mucosal realism	6.0
			Endoscopic view	6.0
			Paradoxical motion	6.0
			Resistance	5.0
			Overall fidelity	5.0

Table 7: Summary of face /content validation evidence for colonoscopy simulation (higher Likert values according to the number of points of the survey scale illustrates increasing evidence of validity)

e) Construct validity of the Simbionix GI mentor II (Simbionix, USA)

There are 4 studies confirming construct validity of the Simbionix GI mentor II simulator (Simbionix, USA) (Koch et al., 2008a, Fayez et al., 2010, Felsher et al., 2005, Grantcharov et al., 2005) for the following common endpoints: procedure time (Fayez et al., 2010, Felsher et al., 2005), efficiency (Fayez et al., 2010, Felsher et al., 2005, Grantcharov et al., 2005), loop formation (Fayez et al., 2010, Felsher et al., 2005, Grantcharov et al., 2005), caecal intubation time (Felsher et al., 2005), polypectomy rate (Felsher et al., 2005) and % mucosa visualised (Grantcharov et al., 2005) (modules 1.1, 1.3, 1.7, 2.1 and 5). One paper did not reference the module that they were assessing (Grantcharov et al., 2005). One study failed to demonstrate construct validity for any endpoints when comparing intermediate, experienced and expert users (Koch et al., 2008a).

f) Construct validity of the Accutouch HT Immersion (Immersion Medical, Germany)

Five studies demonstrated the construct validity of the Accutouch simulator (Immersion Medical, Germany) (Datta et al., 2002, MacDonald et al., 2003, Mahmood and Darzi, 2003, Sedlack and Kolars, 2003) for the following common endpoints: total procedure time (Datta et al., 2002, MacDonald et al., 2003, Mahmood and Darzi, 2003, Sedlack and Kolars, 2003), % mucosa visualised (Datta et al., 2002, Mahmood and Darzi, 2003), polypectomy rate (MacDonald et al., 2003), efficiency of screening (Datta et al., 2002) and caecal intubation time (Sedlack and Kolars, 2003) (*case 1, 3 and 4*). Two studies (MacDonald et al., 2003, Sedlack and Kolars, 2003) failed to demonstrate construct validity for several endpoints when comparing intermediate, experienced users.

g) Construct validity of the Olympus Endo Ts-1 (2nd Generation) (Olympus, Japan)

Two studies (Datta et al., 2002, Koch et al., 2008b, Haycock et al., 2009a) reported the construct validation of the Olympus Endo Ts-1 2nd Generation (Olympus, Japan) simulator for the following common endpoints: caecal intubation, time procedure time, pain and loop formation (Datta et al., 2002, Koch et al., 2008b, Haycock et al., 2009a), (Cases: sigmoid N loop moderate transverse loop +/- low pain threshold, sigmoid α loop moderate transverse loop and sigmoid N loop with γ transverse loop). In one study, (Haycock et al., 2009a) construct validity was not demonstrated for these endpoints in the expert versus intermediate group. There was no evidence for the construct validity of therapeutic measures in these two studies.

h) Evidence for the criterion validity of colonoscopy simulators

There were no reports of criterion validity for any VR simulator. There is one study that demonstrated preliminary criterion validity of the simulator for caecal intubation time (Sedlack, 2011). Authors showed that caecal intubation times on the simulator were comparable to their individual patient-based times obtained from a training database (Sedlack, 2011).

(Study) Simulator	*n*	Groups (previous colons)	Validation	Modules (Mod) / Cases	Task	Endpoints
(Fayez et al., 2010) Simbionix GI mentor II (Simbionix, USA)	20	N (<5 scopes) n=12 E (>50 scopes) n= 8	Construct	Mod 1 (Case 1&7) Mod 2 (Case 1)	Navigation Polypectomy	Time, red out, excessive pressure, %mucosa, %clear view, %pain, looping, %efficiency
(Felsher et al., 2005) Simbionix GI mentor II (Simbionix, USA)	75	Co 1 (n=37) N n=14, E n= 23 Co 2 (n=38) N n=13, E n=25	Construct	Co 1, Mod 1 and 5 Co 2, Mod 2 (case 1)	Navigation Polypectomy	Time to caecum, %mucosa, polypectomy rate % biopsy rate, %time to clear view, time in pain, efficiency ratio
(Grantcharov et al., 2005) Simbionix GI mentor II (Simbionix, USA)	28	N (0 scopes) n=10 I (<50 scopes) n=10 E (>200) n=8	Construct	Single case	Navigation	Time, %mucosa, efficiency, time with clear view, excessive local pressure, pain, time with pain, loop formation
(Koch et al., 2008a) Simbionix GI mentor II (Simbionix, USA)	105	N (0 scopes) n=35 I (<200 scopes) n=15 E (200-1000 scopes) n=20 Expert (EX) (>1000 scopes) n=35	Construct Content	Mod 1 (Case 1 and 3) 1 hand eye test Realism survey	Navigation	Time to caecum, %time with clear view, lost view of lumen, excessive local pressure, %time in pain, loop formation

Table 8 (*part 1*): The validation of the **Simbionix GI mentor II** (Simbionix, USA). Summary of methodology during the validation process. N = novice, I = intermediate, E = experienced, Ex = Expert, n = number of participants in group, co = cohort

(Study) Simulator	*n*	Groups (previous colons)	Validation	Modules/Cases	Task	Endpoints
(Datta et al., 2002) Accutouch HT Immersion (Immersion Medical, Germany)	45	N (0 scopes) I (5-50 scopes) E (>200 scopes)	Construct	Case 1 and Case 4	Navigation	%mucosa visualised, time taken, path length, red-out and efficiency ratio*
(MacDonald et al., 2003) Accutouch HT Immersion (Immersion Medical, Germany)	34	N n=10 I n=19 E n=5	Construct	3 scopes increasing difficulty with pathology to identify	Navigation Identify lesion	Time, insertion length, %bowel visualised, pain, red out , air left in bowel, max force, air levels, %cancers/lesions visualized, %perforations
(Mahmood and Darzi, 2003) Accutouch HT Immersion (Immersion Medical, Germany)	25	N (<10 Scopes) n=11 I (11-100 Scopes) n=7 E (>101 Scopes) n=7	Construct and face	Module 3 or 4	Navigation	Time, %mucosa, path length, mean incidence of perforation
(Sedlack and Kolars, 2003) Accutouch HT Immersion (Immersion Medical, Germany)	22	N (0 scopes) n=6 I (88-202 scopes) n=6 E (482-694 scopes) n=10	Content, Construct and Face	Cases 3 and 4 and Realism survey	Navigation Identify lesion	%mucosa, time, pain, max.depth insertion, air insufflated, air remaining, max.force, %complications, %identification of path
(Moorthy et al., 2004) Accutouch HT Immersion (Immersion Medical, Germany)	20	N (1-10 scopes) n=7 I (20-80 scopes) n=7 E (>200 scopes) n=6	Construct	Case 4, analysis of endoscopic / hand view	Navigation Hand movement	Hand movement, force, endoscopic handing, red out, patient pain, flow of procedure, time taken, insertion,%mucosa

Table 8 (*part 2*): The validation of the **Accutouch HT Immersion** (Immersion Medical, Germany). Summary of methodology of papers that were included in the systematic review, N = novice, I = intermediate, E = experienced, n = number of participants in group

(Study) Simulator	*n*	Groups (previous colons)	Validation	Modules/Cases	Task	Endpoints
(Haycock et al., 2009a) Olympus Endo 2nd Generation (Olympus, Japan)	34	N (0 Scopes) n=10 I (<1000 scopes) n=13 E (>1000 scopes) n=11	Face, content and construct	Realism survey Various loop formations*	Navigation Loop tasks	Caecal intubation, time taken, excess inflation, variable stiffness use, shaft insertion parameters, loop formation and time with loop
(Koch et al., 2008b) Olympus Endo 2nd Generation (Olympus, Japan)	49	N (0 scopes) n=26 E (>1000 scopes) n=23	Face, content and construct	Realism survey, Sigmoid N loop transverse γ loop	Navigation	Dexterity, time to caecum, shaft insertion force, shaft torque, tip section force, max patient pain.
(Sedlack et al., 2007b) Endo X ex-vivo model (Endo-X, USA)	39	N (no scopes) n=13 I (100-150) n=13 E (high volume centres) n=13	Content, Construct and Criterion	X1 simulated colonoscopy Realism survey	Navigation	Caecal intubation, depth insertion, time, scope length to reach caecum, %mucosa, quality of mucosal examination
(Woo et al., 2008a) KAIST-Ewha (KAIST, Korea)	5	E (2500-5000 scopes) n=3 I (1000 scopes)n=2	Face	Realism survey after 3 colonoscopies	Navigation	Ten items to evaluate realism

Table 8 (*part 3*): The validation of the **Olympus Endo 2nd Generation** (Olympus, Japan), **KAIST-Ewha** (KAIST, Korea) and the **Endo X ex-vivo model** (Endo-X, USA). Summary of methodology of papers that were included in the systematic review, N = novice, I = intermediate, E = experienced, n = number of participants in group*Sigmoid N & Moderate transverse loop, Sigmoid N Moderate Transverse loop with low pain tolerance, Sigmoid α Moderate Transverse loop

Sim	Study	Group/ module	Procedure time	Caecal intubation time	Efficiency / efficiency ratio	Loop formation	Patient pain	Time with clear view	% Cancer lesions visualised	% Mucosa visualised	% Polyp/ Bx rate
Simbionix GI mentor II	(Fayez et al., 2010)	N vs E 1.1	<0.01	-	0.08	0.38	0.30	0.61	-	0.13	-
		N vs E 1.7	<0.01	-	<0.01	0.93	0.58	0.05	-	0.01	No data
		N vs E 2.1	0.03	-	<0.01	0.64	0.90	0.11	-	0.07	No data
	(Felsher et al., 2005)	N vs E gp1	-	0.07	0.02	-	0.60	0.05	-	0.02	0.03 / 0.70
		N vs E gp 2	-	0.01	0.03	-	0.30	0.90	-	0.30	0.01
	(Grantcharov et al., 2005)	N vs I vs E	<0.001	-	0.001	<0.001	0.004	0.001	-	0.001	-
	(Koch et al., 2008a)	Ex vs E1.1	-	0.962	-	0.020	0.077	0.621	-	0.621	-
		Ex Vs I 1.1	-	0.141	-	0.018	0.018	0.259	-	0.259	-
		I vs E 1.1	-	0.166	-	0.547	0.385	0.617	-	0.617	-
		I vs N 1.1	-	0.000	-	0.743	0.070	0.177	-	0.177	-
		Ex vs E1.3	-	0.969	-	0.726	0.154	0.297	-	0.297	-
		Ex Vs I 1.3	-	0.326	-	0.090	0.111	0.757	-	0.757	-
		I vs E 1.3	-	0.257	-	0.184	0.771	0.394	-	0.394	-
		I vs N 1.3	-	0.000	-	0.040	0.584	0.104	-	0.104	-

Table 9 (*part 1*): The validation of the **Simbionix GI mentor II** (Simbionix, USA). Statistical outcomes for each assessment parameter analysed by groups for each paper that focussed on the construct validation of the Simbionix GI mentor II (p values were significant if <0.05, indicated by the grey boxes). E = expert, I = Intermediate, N = novice, - = data unavailable or not measured in the study. Bx = Biopsy.

Sim	Study	Group	Procedure time	Caecal intubation time	Efficiency / efficiency ratio	Loop formation	Patient pain	Time with clear view	% Cancer lesions visualised	% Mucosa visualised	Polyp/ Bx rate %
Accutouch HT immersion	(Datta et al., 2002)	N vs I vs E	<0.001	-	<0.001	-	-	-	-	<0.001	-
	(MacDonald et al., 2003)	E vs I	>0.05	-	-	-	>0.05	-	>0.05/>0.05	>0.05	-
		I vs N	>0.05	-	-	-	<0.001	-	<0.001/<0.05	>0.05	-
		E vs N	<0.001	-	-	-	>0.05	-	<0.001/<0.05	>0.05	-
	(Mahmood and Darzi, 2003)	N vs I vs E	0.005	-	-	-	-	-	-	0.001	-
		N vs I vs E	0.030	-	-	-	-	-	-	0.001	-
	(Sedlack and Kolars, 2003)	E vs I	<0.01	<0.01	-	-	>0.05	-	>0.05	>0.05	>0.05
		I vs N	<0.01	<0.01	-	-	>0.05	-	>0.05	>0.05	>0.05
		E vs N	<0.01	<0.01	-	-	>0.05	-	>0.05	>0.05	>0.05
	(Moorthy et al., 2004)	N vs I vs E	0.75	-	-	-	-	-	-	0.21	-

Table 9 (part 2): The validation of the **Accutouch HT Immersion** (Immersion Medical, Germany). Statistical outcomes for each assessment parameter analysed by groups for each paper that focussed on the construct validation of the Accutouch HT immersion (p values were significant if <0.05, indicated by the grey boxes). E = expert, I = Intermediate, N = novice, - = data unavailable or not measured in the study.

Sim	Study	Group	Procedure time	Caecal intubation	Efficiency	Loop formation	Patient pain	Time with clear view	% Cancer lesions	% Mucosa visualised	Polypectomy/ Bx rate %
Olympus 2nd Gen	(Haycock et al., 2009a)	E vs I	1.00	1.00	-	1.00*	NvsIvsE 0.43	-	-	-	-
		I vs N	<0.001	0.003	-	0.05*		-	-	-	-
		E vs N	<0.001	<0.001	-	0.02*		-	-	-	-
	(Koch et al., 2008b)	E vs N	-	0.001	-	-	0.018	-	-	-	-
Endo X	(Sedlack et al., 2007b)	E vs I	0.036	0.001	<0.001	-	-	-	-	0.001	-
		I vs N	<0.001	0.011	<0.001	-	-	-	-	0.002	-
		E vs N	<0.001	0.011	<0.001	-	-	-	-	<0.001	-

Table 9 (part 3): The validation of the **Olympus Endo 2nd Generation** (Olympus, Japan) and the **Endo X ex-vivo model** (Endo-X, USA). Statistical outcomes for each assessment parameter analysed by groups for each paper that focussed on the construct validation of the Olympus 2nd Generation and the Endo X (p values were significant if <0.05, indicated by the grey boxes). E = expert, I = Intermediate, N = novice, - = data unavailable or not measured in the study, * refers to time with sigmoid loop after intubation

2.4 Discussion

a) Summary of review work

The majority of published work included in this review focuses on proving the construct validity of VR simulators. There is some evidence to support the face and content validity of VR and ex-vivo bovine simulators. However, there is no evidence demonstrating the criterion validity of VR colonoscopic simulation and only one paper which reports the criterion validity of an ex-vivo animal tissue model. There is also minimal evidence for the validation of colonic polypectomy simulation.

b) VR validation

The focus on VR simulation can be explained in two ways. Firstly, the endpoints for each simulation are recorded by the instruments' computer system. This standardises the assessment and makes data collection a relatively undemanding process. Secondly, VR modules/cases can create a range of procedures of varying difficulty with the click of a button. This avoids the need for a time consuming "set-up" process which can accompany animal tissue validity testing. The most statistically significant set of values across the all VR assessment parameters were demonstrated in the paper by Grantcharov et al., 2005. This may have been due to the fact statistical comparisons were made across 3 vastly different groups of expertise whilst the majority of other studies compared results on a group-by-group basis.

c) The limitations of VR colonoscopy simulation

Although the currently available computer simulation scenarios are very reproducible, it can be argued that they lack the complexity and fidelity to be useful in any meaningful way (Sedlack et al., 2007b). This review shows that VR simulators are often unable to

significantly distinguish between intermediate level trainees and experts. This may reflect that VR at present may only be useful for teaching basic endoscopic skills to the novice trainee. It may be that the level of difficulty needs to be increased in order to make the distinction between more experienced groups. This is a view shared by several authors (Hassan et al., 2008, Maschuw et al., 2010, Maagaard et al., 2011). An alternative option is the composite animal model. Sedlack *et al* 2007, has shown the ability of the bovine model to discriminate between senior trainees and experts for several endpoints (Sedlack et al., 2007b). This coupled with favourable realism scores and difficulty ratings creates an area of interest for future research. Composite models also use "real" endoscopes and accessories including electrosurgery so add an extra level of realism that may be useful in the clinical environment.

Several studies in this review reported endpoints that were unable to distinguish between levels of experience in a statistically significant way. This questions the construct validity of these endpoints. Researchers in the past have suggested that VR endpoints are, with few exceptions invalid as meaningful metrics (Sedlack and Kolars, 2003) (Aabakken et al., 2000). From this review, we can report that the most valid VR assessment parameters across all studies are: total procedure time, caecal intubation time, efficiency and % mucosa visualised. These could be considered as the most reliable indicators of performance in future VR work[‡].

d) Limitations of the review

This review was limited by the degree of variation in methodology between included studies. This was particularly true for group assignment when examining construct validity. Participants were assigned to groups according to the number of previous colonoscopies performed. This raises two issues. Firstly, it could be argued that the number of previous colonoscopies is not a good measure of experience. In effect, using this parameter may

[‡] These valid assessment metrices are utilised in Chapter 5 of this thesis

influence the reliability of any validation study. A more appropriate measure of experience could be provided by reviewing scores for previously completed DOPS assessments. Secondly, the definition of "novice", "intermediate" or "experience" operators was highly variable between studies. For example, one study (Moorthy et al., 2004) considered participants with >200 previous colonoscopies as "experienced" whereas another (Koch et al., 2008b) considered >1000 to be an appropriate figure. This may not have adversely influenced construct validity within a study, but reduced the reliability of inter-study analysis.

e) Papers published following review work

From January 2012 to July 2013, the above systematic review search strategy was repeated in order to identify further validation studies in colonoscopy simulation. Two further papers were subsequently identified (**Table 10** and **Table 11)**. Plooy *et al* demonstrate the construct validity of the Kyoto Kagaku Colonoscope Bench Model (Kyoto Kagaku Ltd, Japan) (Plooy et al., 2012). This paper shows that experienced colonoscopists (n=21) had significantly higher completion rates and shorter time to caecum for all simulated cases when compared to novices (n=18) (Plooy et al., 2012).

McConnell *et al* evaluated the content and criterion validity of the CAE Endoscopy VR simulator (CAE, USA) focussing on both colonoscopy and EGD modules as predictors of performance (McConnell et al., 2012). They showed statistically significant differences between 5 novices and 6 experts in only 19 of 57 (33%) performance metrics. The statistically significant parameters were time-related.

(Study) Simulator	*n*	Groups (No. of colonoscopies)	Validation	Modules/ Cases	Task	Endpoints
(Plooy et al., 2012) Kyoto Kagaku Colonoscope Bench Model (Kyoto Kagaku Ltd, Japan)	39	Novices (n=18) Experts (n=21)	Construct	Not applicable	2 attempts at each of 4 standard cases	Completion rates, time to caecum and peak forces applied to the colon model
(McConnell et al., 2012) CAE Endoscopy VR simulator (CAE, USA)	11	Novices (n=5) (year 1, fellows) Experts (n=6) (Attending physicians)	Content and Criterion	Diagnostic colonoscopy, colon biopsy, colonic polypectomy and OGD with 6 varying degrees of difficulty	18 simulated colonoscopies and 6 simulated OGD's	Realism Survey and 57 objective parameters measured by the simulator including total procedure time, time spent in contact with mucosa and time to anatomical landmarks

Table 10: The validation of the **Kyoto Kagaku Colonoscope Bench Model** (Kyoto Kagaku Ltd, Japan) and the **CAE Endoscopy VR simulator** (CAE, USA). Summary of methodology of studies published following systematic review

Sim Type	Study	Group/ module	Procedure time	Caecal intubation time	Efficiency / efficiency ratio	Loop formation	Patient pain/ peak force	Time with clear view	% Cancer lesions visualised	% Mucosa visualised	Polyp/ Bx rate %
Kyoto Kagaku Colonoscope Bench Model (Kyoto Kagaku Ltd, Japan)	(Plooy et al., 2012)	N vs E	<0.01	<0.0001	-	0.01	<0.001	-	-	-	-
CAE Endoscopy VR simulator (CAE, USA)	(McConnell et al., 2012)	N vs E	<0.001	<0.001	-	-	0.018	<0.001	-	-	-

Table 11: The validation of the **Kyoto Kagaku Colonoscope Bench Model** (Kyoto Kagaku Ltd, Japan) and the **CAE Endoscopy VR simulator** (CAE, USA). Statistical outcomes for each assessment parameter analysed by groups for each paper that focussed on the construct validation of the CAE Endoscopy VR simulator and Kyoto Kagaku Colonoscope Training Model, N = novice, E = expert, - = data unavailable or not measured in the study (p values were significant if <0.05, indicated by the grey boxes).

f) Conclusions

There is a plethora of validation evidence for VR colonoscopy simulation. This focuses on, construct, content and face validity for specific modules and endpoints. To fully validate VR models, more study of criterion validity is needed. Further validation studies may be required to evaluate each individual case/module within VR simulators using relevant endpoints. Standardised methodology between studies would provide stronger evidence. Preliminary results for animal composite colonoscopy simulation are encouraging for content, construct and criterion validity. Given their ability to clearly distinguish between intermediate and experienced users they may have the advantage of providing a useful training tool for senior endoscopic trainees. Further research on therapeutic colonoscopy simulation is needed if these tools are to be used to assess progression into independent practice.

g) Key results from systematic review

- The current scientific literature focuses on evaluating the construct validity of VR colonoscopy simulators
- VR colonoscopy simulators lack the ability to distinguish between intermediates and experts
- The assessment parameters for VR colonoscopy simulation may lack clinical relevance
- Only one report of construct / criterion validity for an ex-vivo colonoscopy simulator exists and there is no evidence that tests the validity of an ex-vivo polypectomy model

Chapter 3: Materials and Methods

Constructing the WIMAT colonoscopy suitcase

3.1 The WIMAT colonoscopy suitcase

The WIMAT colonoscopy suitcase is an ex-vivo porcine simulator designed to teach the principles of the colonic polypectomy. This model has been designed at the Welsh Institute for Minimal Access Therapy (WIMAT) in response to a lack of existing commercial simulators that focus on colonoscopic polypectomy training.

a) Simulator capacity

This simulator has the potential to simulate a range of polypectomy procedures. Different types of Polypoid (0-1p and 0-1s) (**Figure 16A**) and Non Polypoid / sessile (0-11a) (**Figure 15B**) lesions can be created with the potential for simulated haemorrhagic scenarios (**Figure 15C**). The simulator can be used with any endoscope and endoscopic accessories[§].

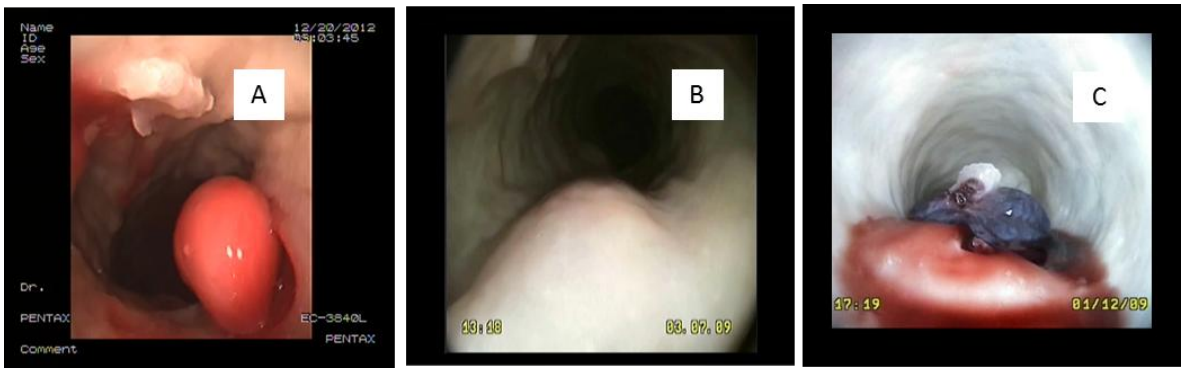


Figure 15: The WIMAT colonoscopy suitcase. Endoscopic view of simulated polypoid lesions A = 0-1p polyp, B = 0-1s polyp and C = Bleeding pedunculated polyp

b) Outer casing

Porcine rectum, used to replicate human bowel, is housed in a portable polymer suitcase (Storm Case IM 2600, Hardigg®) with a hole made in one end to simulate the anus (**Figure 17 A - B**). This hole is cannulated with a 15mm Ethicon XL port (Johnson and Johnson,

[§] The attached DVD demonstrates the set-up and capacity of the suitcase model

USA) and secured internally with a plastic ring clip. This allows the passage of the colonoscope into the suitcase and provides an air tight seal to enable insufflation of the colon.

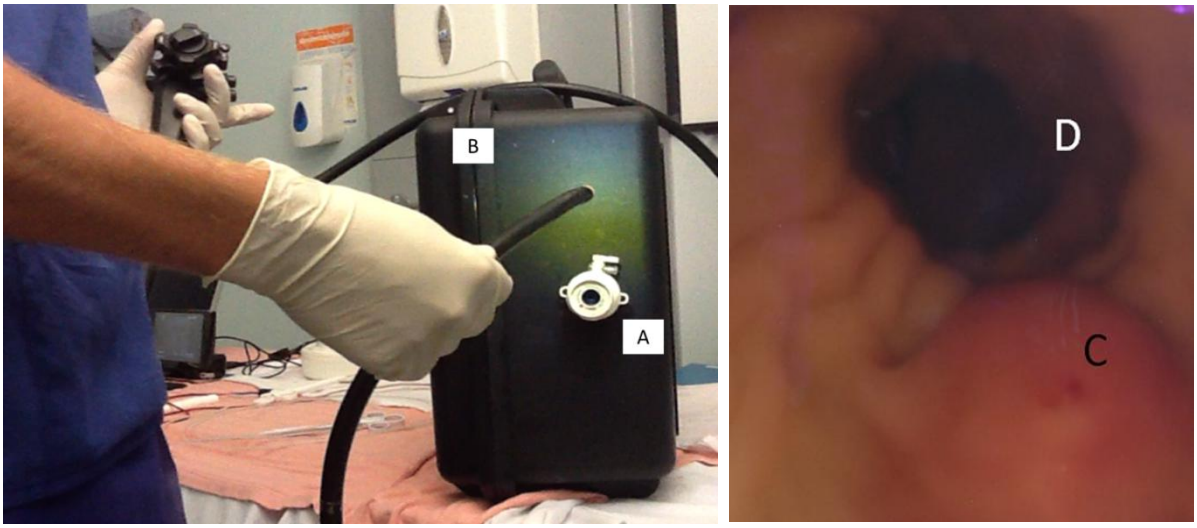


Figure 16: External aspect of the WIMAT colonoscopy suitcase. A = 15mm Ethicon XL port (Johnson and Johnson, USA) for endoscopic cannulation, B = Portable polymer suitcase (Storm Case Im2600, Hardigg®, UK), C = simulated sessile polyp and D = luminal fold

Inside the suitcase is a removable metal mesh base (**Figure 18**). This accommodates a crocodile clip which is connected to an electrosurgery unit for use during the simulated polypectomy. Foam segmentors are mounted onto metal rails (**Figure 17**) which are secured with wing nuts onto the mesh base. The porcine bowel is then passed through the inside of the foam segmentors, a curved piece of 55mm standard exhaust piping (to simulate the sigmoid colon) and then through a second set of segmentors. The anal end of the specimen is then attached to the Ethicon XL port (Johnson and Johnson, USA) using cable ties. When the bowel is inflated, the foam segmentors indent on the serosal aspect of the bowel (to represent haustral folds) to give it a realistic luminal appearance (**Figure 16 D**).

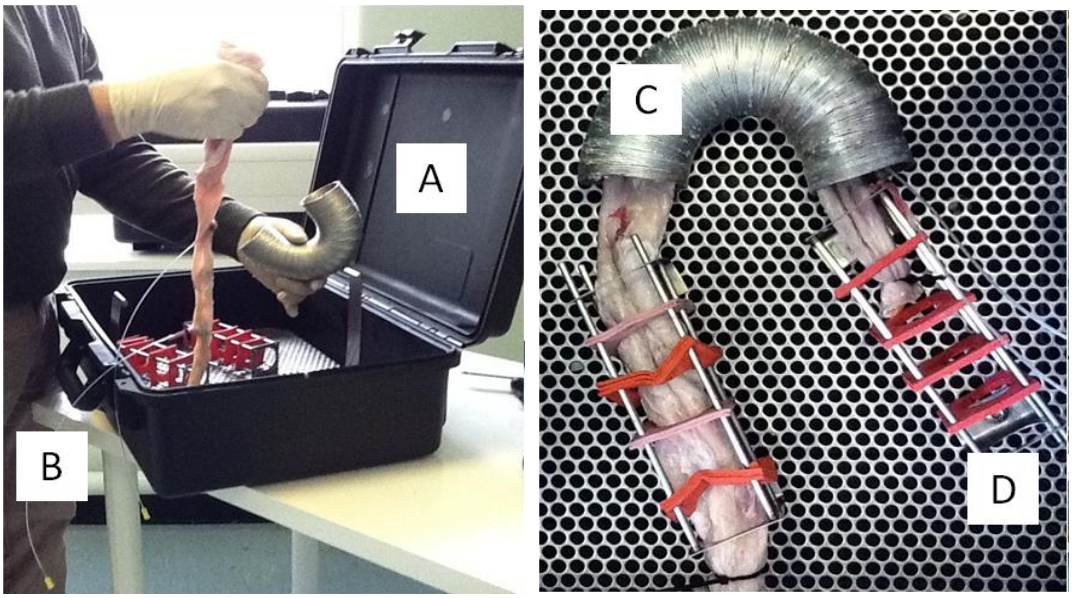


Figure 17: The WIMAT colonoscopy suitcase external view (Left) and internal view (Right). A = outer casing B = cannulated polyps for simulated bleeding C = exhaust piping D = foam segmentors and metal rail construct

c) Porcine specimens

These specimens originate from low risk category 3 Animal By-Products (ABP). All animal samples are handled and disposed of according to strict internal protocol. Each specimen undergoes a standardised cleaning process. These are defrosted and everted to expose the mucosa. The mucus from the inner bowel is then removed.

d) Polyp construction

Three types of polyps are constructed. These included sessile, pedunculated (non-bleeding) and pedunculated (bleeding) polyps. Sessile polyps are created by injecting a standardised volume of a polyp mix into the bowel submucosal layer (**Figure 19**). The polyp mix is a solution which solidifies at room temperature and does not breakdown when re-frozen. This is made by combining 2mls of red colouring agent, 100g of gelatine, 10g of Suet and Oil mix,

300mls of water, 5mls of washing detergent and 10g of salt. The mix was warmed using a T.ARE heating magnetic stirrer (VELP®) to a temperature of 90ºC. Once liquefied, it is injected into the submucosa, cooling in situ and sealing its position as a sessile polyp.

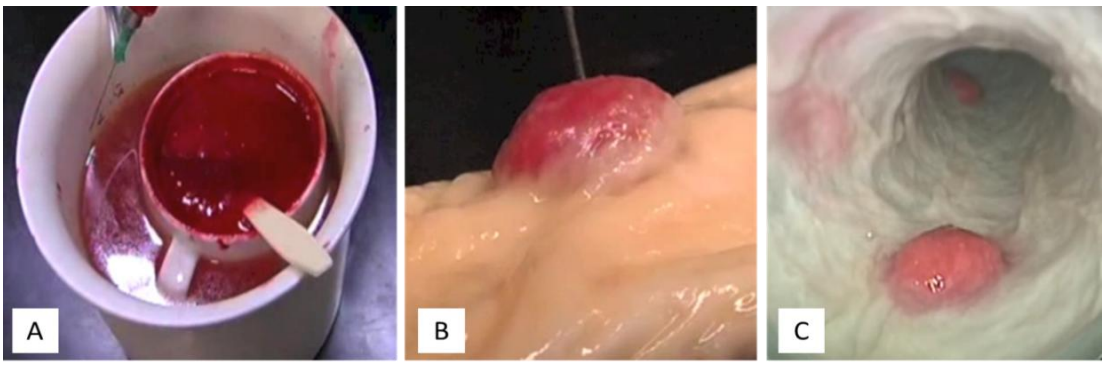


Figure 18: Sessile polyp construction A = Polyp mix, B = injection of polyp mix into submucosa, C = endoscopic view of simulated sessile polyp

Pedunculated (non-bleeding) polyps are constructed by using a 5cm length of sausage skin sealed at one end with a surgeon's reef knot (**Figure 20 A-B**). This is then turned inside out to internalise the knot (**Figure 19 C**), filled with 2mls of liquid polyp mix and allowed to solidify (**Figure 19 D-G**). For pedunculated (bleeding) polyps, this process is repeated and a length of porcine ureter is attached inside each polyp which is then cannulated with a plastic catheter. This catheter is attached to a 50ml Luer-Lock syringe containing simulated blood.

e) Attaching the polyp to porcine bowel

The mucosa of the everted bowel is pierced using a Roberts artery forceps; the stalk of the polyp is grasped and pulled through (**Figure 21A**). The stalk is then secured using a surgeon's reef knot and glue (**Figure 20B and 20C**). The whole specimen is then inverted so that the polyps are transferred to the internal luminal aspect of the lumen of the bowel. **Figure 22** demonstrates an overview of the porcine specimen before and after this inversion process.

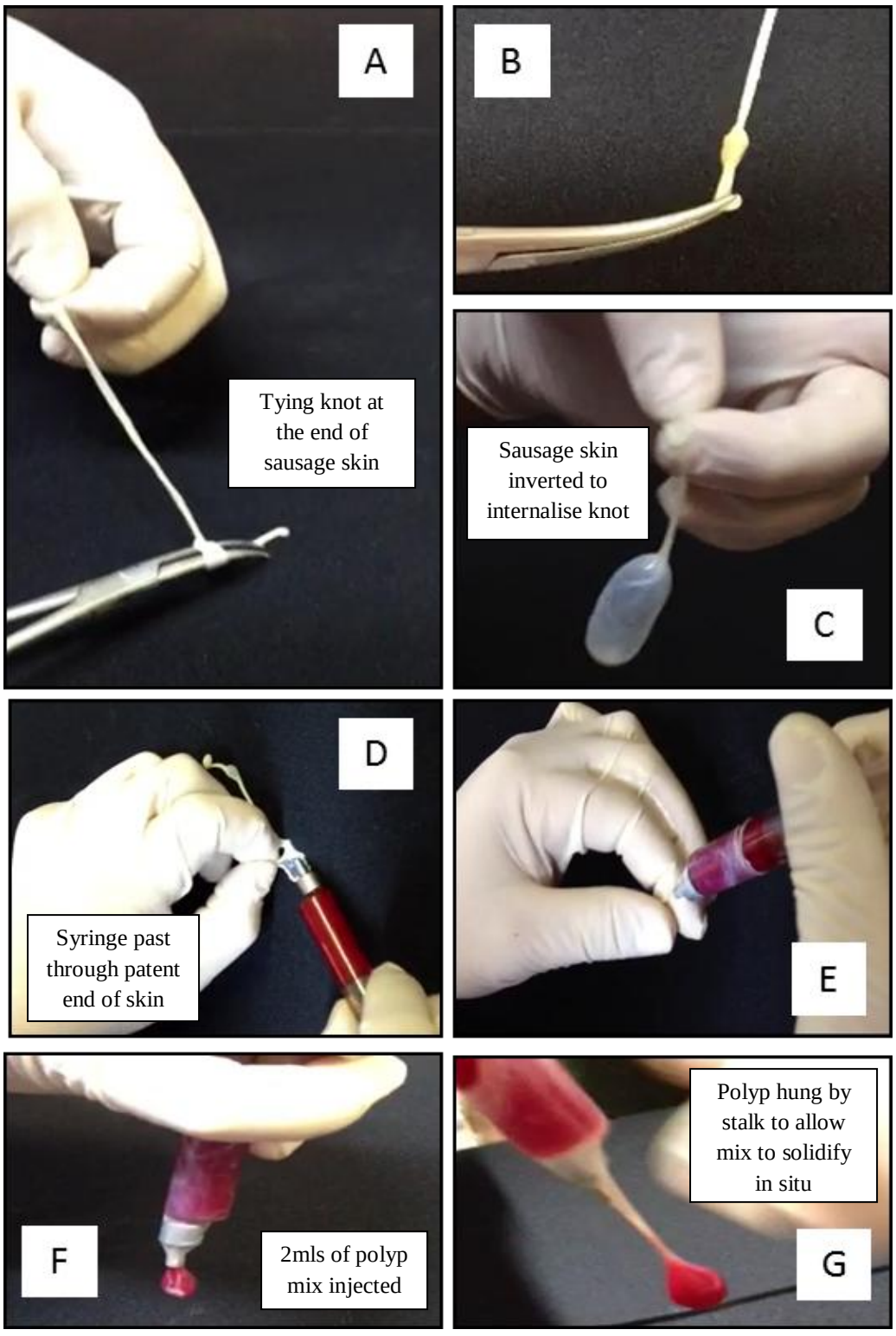


Figure 19: Pedunculated polyp construction (non-bleeding). A and B = securing the distal porcine sausage skin, C = Internalising knot, D-G = Injection of polyp mix

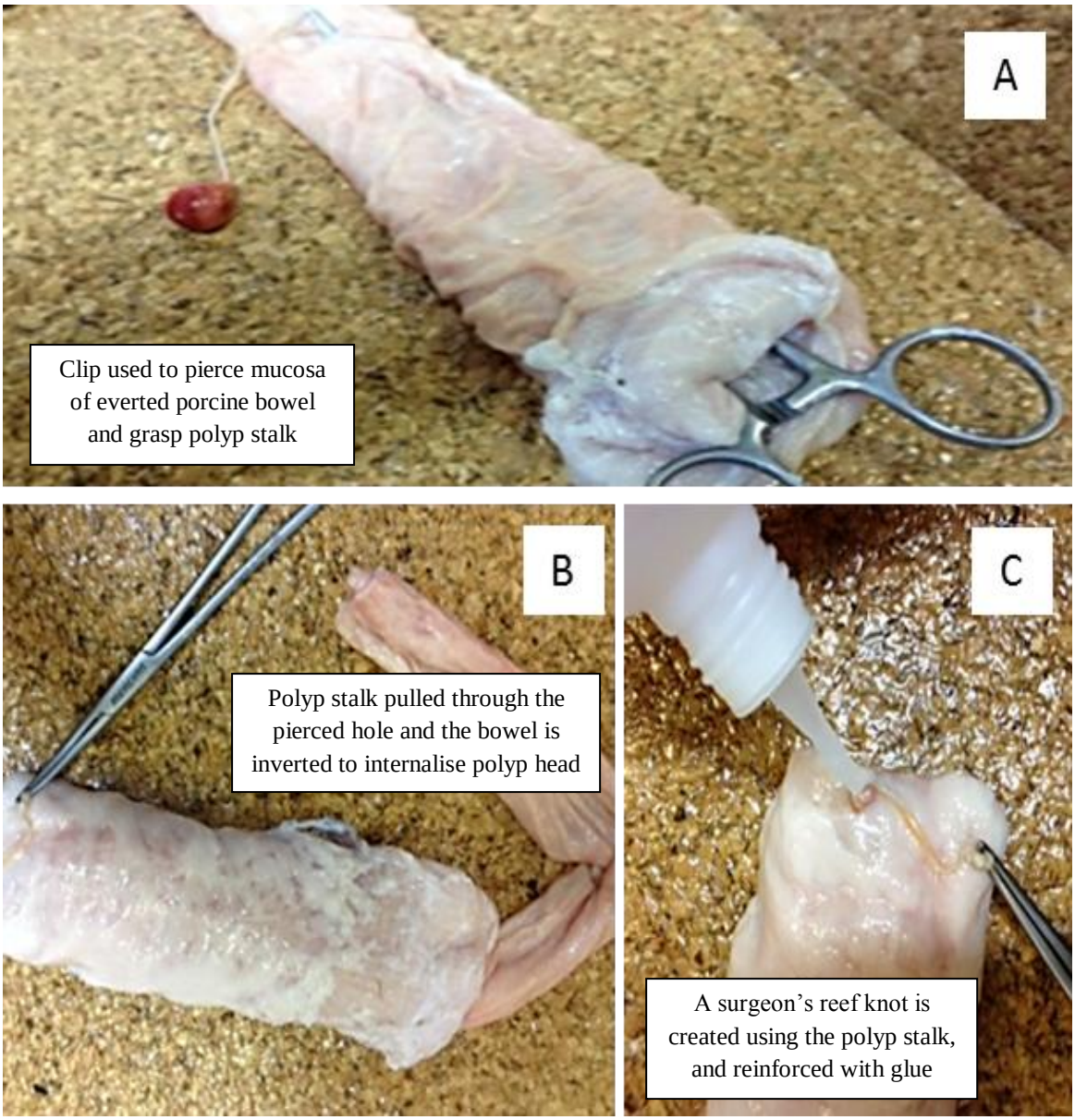


Figure 20: Securing the polyp stalk to porcine bowel wall. A = Piercing the everted bowel mucosa and grasping the polyp stalk, B and C = Anchoring the polyp stalk to the serosal aspect of the bowel

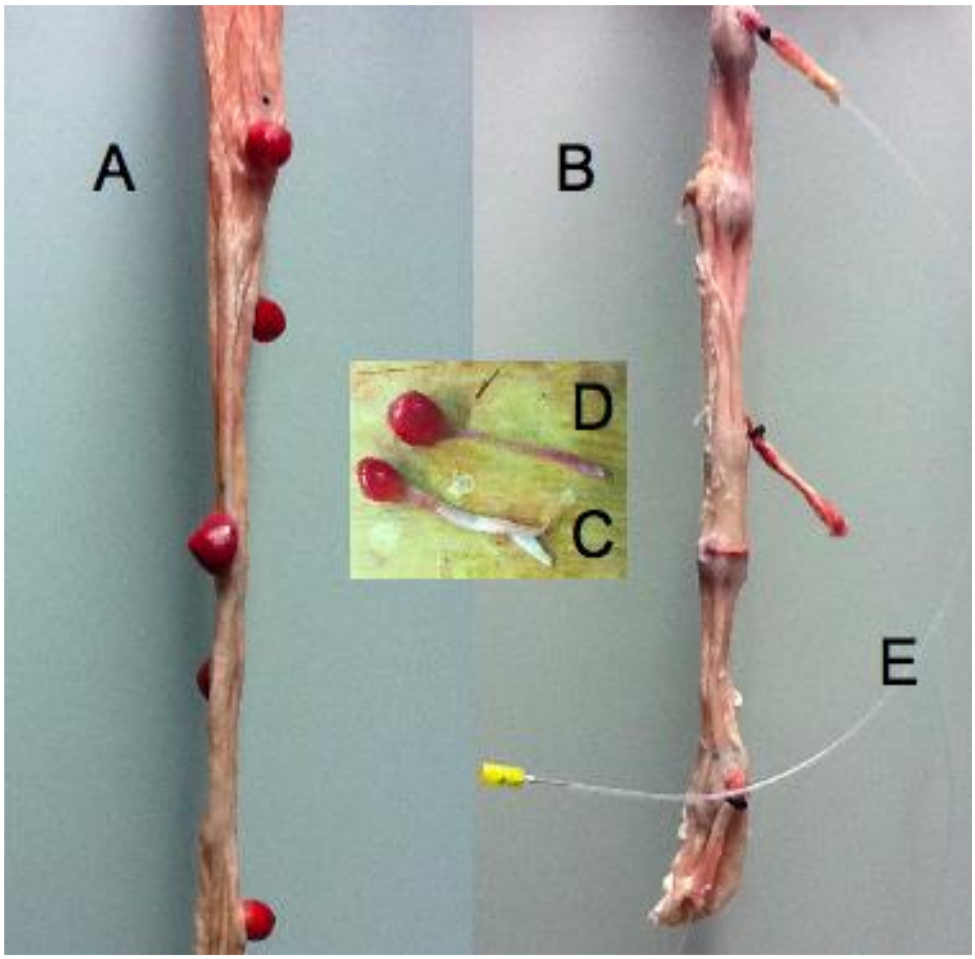


Figure 21: Porcine tissue. A = Bowel everted (mucosa on outside, pedunculated polyps attached), B = Bowel inverted (mucosa on inside), C = Pedunculated (bleeding) polyp, D = Pedunculated (non-bleeding) polyp, E = Catheter inserted into pedunculated polyp to simulate bleeding

f) Endoscopic equipment

One of the main benefits of using the simulator is that any "real-life" endoscopic equipment can be used. This is in contrast to VR simulators where only simulated accessories are available. The Pentax EC-3840L Colonoscope (Pentax Medical, UK), Pentax PVM-2053MD monitor (Pentax Medical, UK) and Pentax EPM-3500 processor (Pentax Medical, UK) were used in this work (**Figure 23**). Standardised endoscopic snares (maximum diameter 25mm) (Cook®, Wilson Cook Medical GI Endoscopy, USA) and 3.0cm Roth Nets® (US endoscopy)

were used for polypectomy and retrieval (**Figure 24**). The Valley Lab™ EZc (Covidien, USA) energy source is placed on a coagulation setting of 40 watts at all times (**Figure 23**). An Eschmann VP25 portable suction unit set at a standard level is also used.

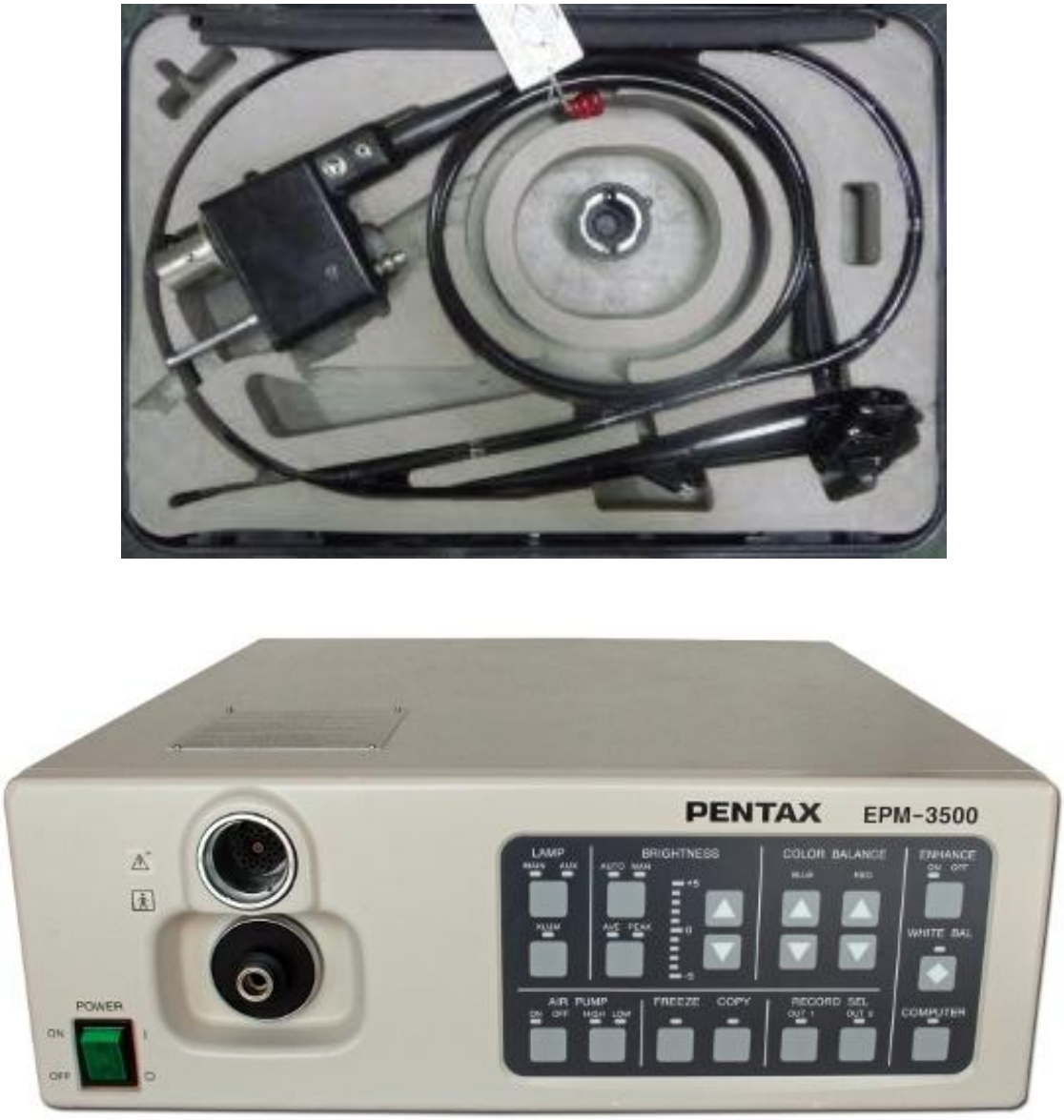


Figure 22: Above = Pentax EC-3840L Colonoscope (Pentax Medical, UK) **a**nd below = Pentax EPM-3500 processor (Pentax Medical, UK)

Figure 23: Top = Endoscopic 25mm snare, Cook® (Wilson Cook Medical, USA), middle = 3.0cm Roth Net® (US endoscopy, USA) and bottom = Valley Lab™ EZc (Covidien, USA)

3.2 Adapting the WIMAT colonoscopy suitcase to test construct and concurrent validity

An important part of simulator design is ensuring that the model can be adjusted for varying degrees of difficulty. This allows users of increasing expertise to benefit from simulation training. Gupta *et al* have recently described a method for determining the difficulty of colonoscopic polypectomy (Gupta et al., 2013). The authors describe a scoring system based on size, morphology, site and access (Gupta et al., 2013). Four polyp levels (with increasing levels of complexity) were identified base on the range of scores obtained (Gupta et al., 2013) (**Table 12**).

a) Defining the technically difficult polyp

Some polyps, due to their size, location, or configuration, are considered more technically challenging and are associated with an increased risk of complications (Gallegos-Orozco and Gurudu, 2010). Polyps that are larger than 15 mm, have a large pedicle, and are flat and/or laterally spreading, are difficult to see or are located in the caecum or any angulated portion of the colon are always considered difficult (Monkemuller et al., 2008, Monkemuller et al., 2009, Saito et al., 2010, Ahmad et al., 2002).

b) Polyp diameter

Polyp size has a significant impact on the difficulty of the polypectomy. In general, sessile or pedunculated polyps >2 cm in diameter are considered difficult. Any polyps >3 cm in diameter represent the most challenging (Gallegos-Orozco and Gurudu, 2010). Most endoscopists need to be adequately trained and equipped to excise the majority of polyps found in a routine colonoscopy which are frequently less than ≤1 cm in diameter (Gallegos-Orozco and Gurudu, 2010). Lesions larger than 1.5cm should be resected using adjunctive

techniques such as submucosal cushion or piecemeal methods (Ahmad et al., 2002, Monkemuller et al., 2008, Saito et al., 2010).

Parameter	Range	Score
Size	<1cm	1
	1-1.9cm	3
	2-2.0cm	5
	3-3.9cm	7
	>4cm	9
Morphology	Pedunculated	1
	Sessile	2
	Flat	3
Site	Left	1
	Right	2
Access	Easy	1
	Difficelult	3
Polyp Level	Range of scores	
Level 1	4-5	
Level 2	6-8	
Level 3	9-12	
Level 4	>12	

Table 12: Scoring system for determining the difficulty level of a polyp (Gupta et al., 2013)

c) Polyp stalk

Pedunculated polyps are generally well suited for simple snare polypectomy however, when the stalk is thick or long (**Figure 25**), polypectomy may be technically challenging or may have an increased risk of immediate or delayed bleeding (Vormbrock and Monkemuller, 2012). In general, polyps with the longest stalks tend to be located within the left colon, as the pedicles are formed by mucosa and submucosa pulled toward the lumen by the peristaltic action of the colon (Vormbrock and Monkemuller, 2012).

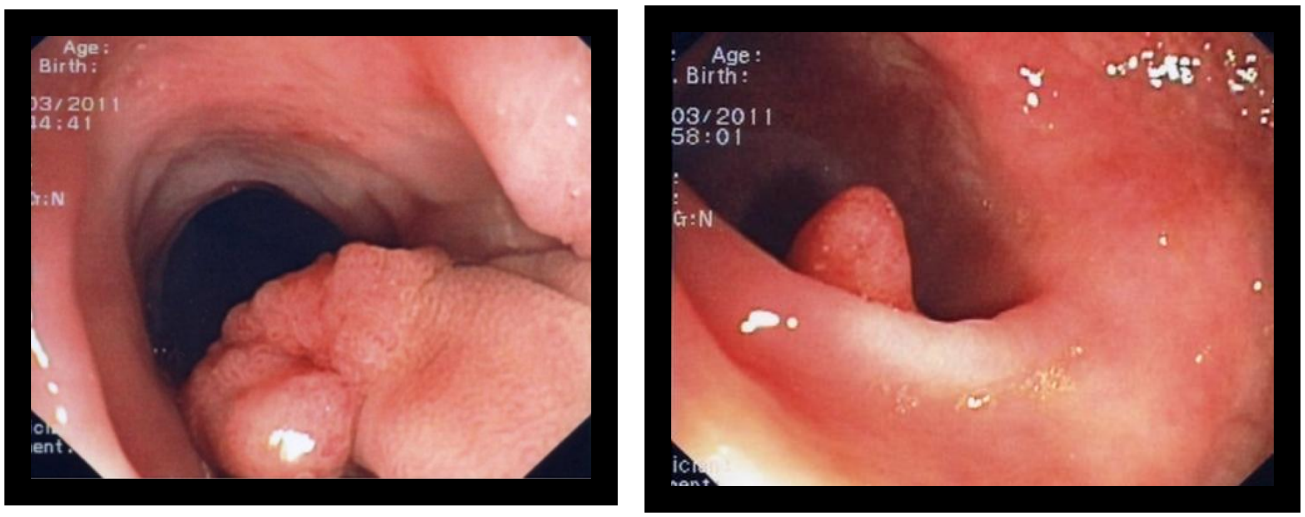


Figure 24: Left = long thickened stalked polyp (Vormbrock and Monkemuller, 2012) and right = polyp located "behind" a luminal fold (Vormbrock and Monkemuller, 2012)

d) Polyp location

Another factor that makes a polyp difficult to treat endoscopically is its location. This can include; in a wall that is difficult to access with a colonoscope; within an area of severe diverticulosis; being behind or wrapped around a fold in a clam-shell fashion (**Figure 24)** occupying more than one third of the colonic circumference; or crossing over 2 haustral folds (Waye, 2005, Tholoor, 2013). Polyps that cross over two haustral folds present two distinct problems (Kudo and Kashida, 2005, Higaki et al., 2003). Firstly, the portion that lies in the valley between two interhaustral septae can be difficult to access (Higaki et al., 2003). Secondly, there is a risk of catching the full thickness of colonic wall in the snare, especially over the haustral fold, inadvertently leading to perforation (Higaki et al., 2003).

e) Creating simulated polyps with different degrees of difficulty

Using the information above we have created 2 polyps to provide 2 hypothesised degrees of difficulty (**Table 12**). As this is primarily a research model for teaching the skills required to perform polypectomy, it was decided that the degree of difficulty needed to be standardised.

Therefore, we have kept the size, morphology, (**Figure 26**) and stalk of the polyp the same and have varied the location.

f) Varying the polyp location in the WIMAT colonoscopy suitcase

Polyp A (simple) has been inserted 25cm from the anal verge in the six o'clock position in front of a luminal fold (**Figures 27 and 28**). This polyp is placed on the "up-going" part of bowel inside a lumen that is 500mm in diameter. It was hypothesised that this position would be optimal for effective curvature of the colonoscope and would allow participants to demonstrate a basic level of procedural skill. Polyp B (complex) has been inserted at 43cm from the anal verge, in a 1 o' clock position, behind a luminal fold, distal to a simulated colonic bend within a luminal diameter of 350mm (**Figures 26 and 27**). This was designed to increase the difficulty of the polypectomy with the aim of differentiating between more experienced users. In reality, polyps can be placed at multiple sites along the length of the porcine bowel.

Feature	Polyp A (simple)	Polyp B (complex)
Morphology	Polypoid lesions: 0-1p	
Size	Diameter = 10mm, Polyp length = 15mm, Stalk = 5mm	
Luminal Diameter	500mm	350mm
Location	25cm from anal verge	47cm from anal verge
	6 o' clock position,	1 o' clock position
	In front of a luminal fold	Behind a luminal fold
	"Up-going" part of bowel	Distal to a colonic bend

Table 13: Features of simulated Polyp A (simple) and Polyp B (complex)

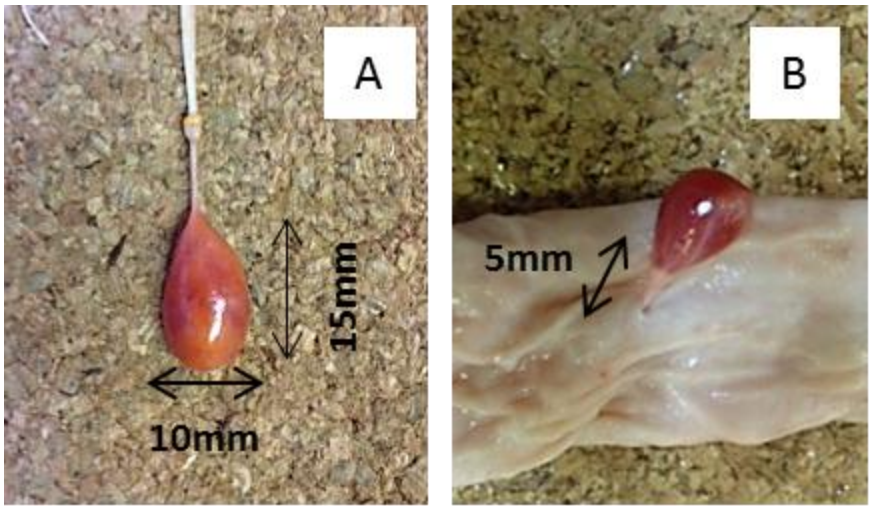


Figure 25: Standardised dimensions of the polyps used for testing construct validity of the WIMAT colonoscopy. A = Length and width of polyp head, B = Length of polyp stalk

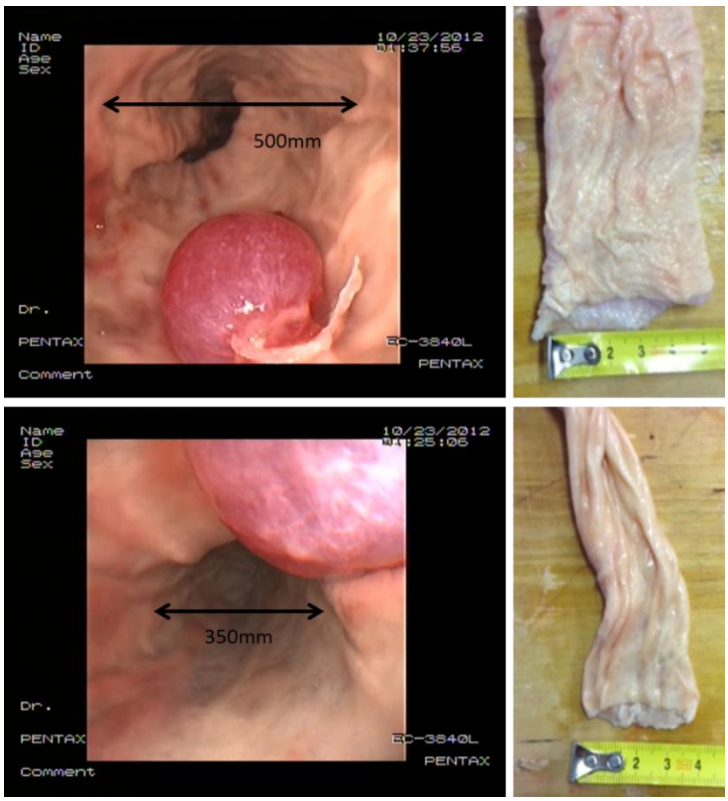


Figure 26: Construct validity - diameter of lumen for polyp A (above) and polyp B (below)

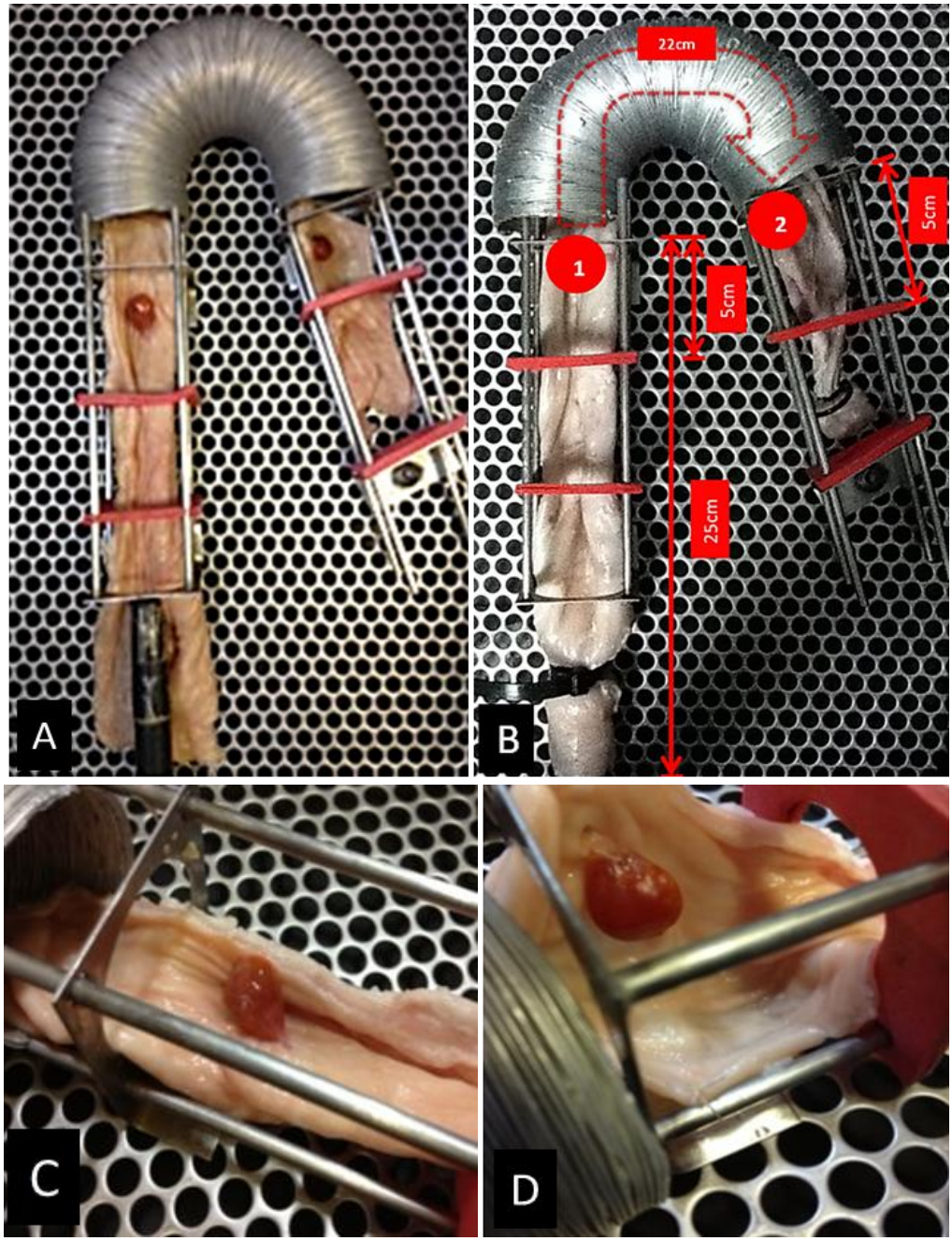


Figure 27: Positioning of polyps for testing construct and concurrent validity. A = internal view of the simulator with anterior aspect of the bowel exposed to demonstrate positions of the two polyps, B = Positional dimensions of the polyps (where, 1 = position of polyp A and 2 = position of polyp B). C = position of polyp A and D = position of polyp B when bowel opened up to expose internal surface

Chapter 4: Validation

Testing the content validity of the WIMAT colonoscopy suitcase

4.1 Measuring the content validity of the WIMAT colonoscopy suitcase: Introduction

a) Content validity

Content validity is the first stage of any simulation validation process (**Table 5**). This is "*the extent to which a measure reflects the trait or domain it purports to measure*" (Sedlack, 2011). Content validity is frequently measured using questionnaires for expert users to measure the haptic and visual realism of a simulator (**Table 5**). The aim of this study was to test the content validity of the WIMAT colonoscopy suitcase.

4.2 Methodology

All participants were recruited from regional and national endoscopic training courses where the simulator was being demonstrated. These were all faculty members, all highly experienced in colonic polypectomy. Participants were either independent Consultant practitioners and senior trainees or nurse endoscopists that were JAG accredited to perform colonoscopy. Each participant completed a snare polypectomy on a simulated pedunculated (bleeding/ non-bleeding) and sessile polyp. All polyps, equipment and endoscopic equipment were standardised throughout. Following the simulation, each participant completed a pre-designed realism questionnaire (**Figure 29**)

a) Construction of an expert realism survey to test content validity

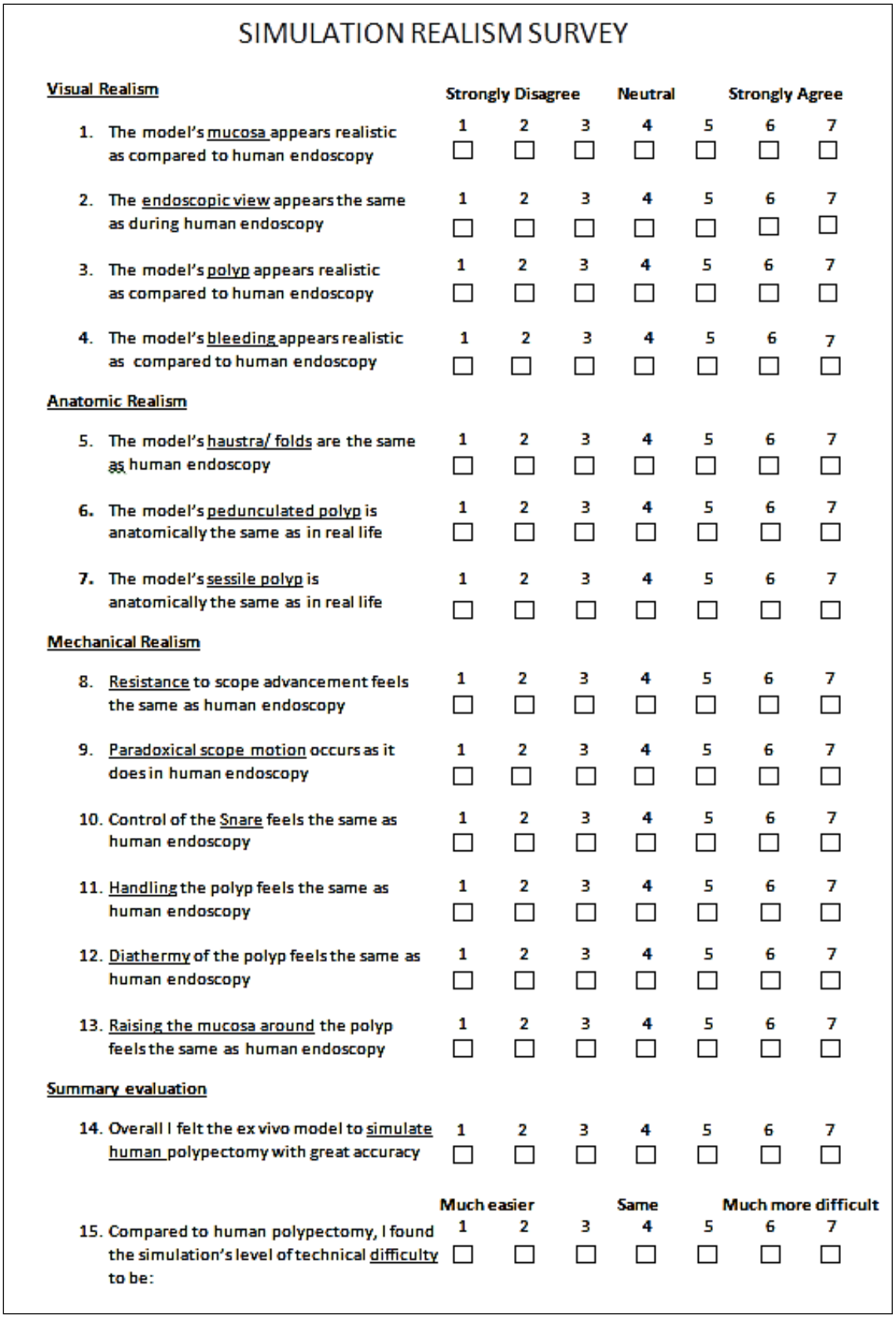
The systematic review (chapter 2) demonstrated that there is only one study which reports the content validity of an ex-vivo colonoscopy simulator (Sedlack et al., 2007b). In this study, Sedlack *et al* focus on the complete validation of a bovine model for colonoscopy training (Sedlack et al., 2007b). A Mayo Clinic Survey (Mayo Clinic, Rochester, MN) was used to grade different aspects of the bovine model's realism and complexity compared to a live

human colonoscopy (**Figure 30**). This validated survey was adapted to evaluate the content validity of the WIMAT colonoscopy suitcase (**Figure 28**). Questions 1-13 of the survey were divided into the following three areas: visual realism, anatomical realism, mechanical realism. Individual parameters within each of the 4 sections were scored on a 7-point Likert Scale (1 = strongly disagree, 4 = neutral, 7 = strongly agree). Question 14 focussed on the overall degree of similarity between the simulated polypectomy and "real life" polypectomy (1 = strongly disagree, 4 = neutral, 7 = strongly agree). Question 15 compared the technical difficulty of human polypectomy with the simulation (1 = much easier, 4 = same, 7 = much more difficult). The full questionnaire used during the study can be found in **Appendix 2a**.

b) Data analysis and power calculations

Assuming that the 7-point scale has a standard deviation of 1.0, it was calculated that 17 participants would give >90% power to detect a difference of 1 point or more on the survey against a hypothetical mean of 4. Realism surveys were analysed using the Wilcoxon signed-rank test on a PASW Statistics 18 (SPSS) for non-parametric data. Median values from the 7-point scale were compared with a hypothetical mean of 4 to determine statistical significance. Demographic data is presented as mean values with 95% confidence intervals[**].

[**] Statistical analysis has been reviewed and approved by Professor P Price



SIMULATION REALISM SURVEY

	Strongly Disagree			Neutral			Strongly Agree
Visual Realism							
1. The model's mucosa appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
2. The endoscopic view appears the same as during human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
3. The model's polyp appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
4. The model's bleeding appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Anatomic Realism							
5. The model's haustra/ folds are the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
6. The model's pedunculated polyp is anatomically the same as in real life	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
7. The model's sessile polyp is anatomically the same as in real life	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Mechanical Realism							
8. Resistance to scope advancement feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
9. Paradoxical scope motion occurs as it does in human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
10. Control of the Snare feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
11. Handling the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
12. Diathermy of the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
13. Raising the mucosa around the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Summary evaluation							
14. Overall I felt the ex vivo model to simulate human polypectomy with great accuracy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
	Much easier			**Same**			**Much more difficult**
15. Compared to human polypectomy, I found the simulation's level of technical difficulty to be:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Figure 28: Questionnaire used for evaluating the content validity of the WIMAT colonoscopy suitcase adapted from (Sedlack et al., 2007b)

Simulation Realism Survey

		Strongly Disagree			Neutral			Strongly Agree
Visual Realism:								
1.	The model's mucosa appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
2.	The endoscopic view appears the same as during human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Anatomic Realism:								
3.	The model's colon length is the same as human endoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
4.	The models haustra/folds are the same as human endoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
5.	The model's degree of angulation in the sigmoid is the same as human colonoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
6.	The model's degree of angulation at the splenic flexure is the same as human colonoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
7.	The model's degree of angulation at the hepatic flexure is the same as human colonoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Mechanical Realism:								
8.	Resistance to scope advancement feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
9.	Paradoxical scope motion occurs as it does in human endoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
10.	Resistance and paradoxical motion occur at the same landmarks as those that occur human endoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
11.	Looping appears to occurs in the same frequency and degree as in human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
12.	Loop reduction appears responds to the same reduction techniques that are used in human endoscopy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Summary Evaluation:								
13.	Overall I felt the *ex vivo* bovine model to simulate human endoscopy with great accuracy:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
		Much Easier			Same			Much More Difficult
14.	**Compared to human endoscopy**, I found the simulation's level of technical difficulty to be:	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐

Figure 29: Mayo Clinic Survey (Mayo Clinic, Rochester, MN) (Sedlack et al., 2007b)

4.3 Results[††]

a) General Results

A total of 17 participants (Male: Female ratio 14:3) completed the questionnaire; 15 (88.2%) Gastroenterologists, 1 (5.9%) Colorectal Surgeon and 1 (5.9%) experienced endoscopic Nurse Specialist. Of the Gastroenterologists, 7 (46.7%) were Consultants, 8 (53.3%) were ST6-7 level or Post Certificate of Completion of Training (CCT) fellows (**Table 13**). All participants were experienced in performing colonoscopy, polypectomy and biopsy. The mean numbers of previous procedures performed by the cohort were 371 (179-689) colonoscopies, 156 (35-355) polypectomies and 165 (42-360) biopsies (**Table 14**). The majority of the cohort had previous experience of using several different polypectomy simulators (**Table 14**).

b) Results of the realism survey

The WIMAT colonoscopy suitcase demonstrated a good level of realism scoring across a range of parameters. The highest scores were for "mucosal realism" (Median score 6.0, p=0.001), "endoscopic snare control" (Median score 6.0, p=0.001), "handling the polyp" (Median score 6, p=0.001) and "raising mucosa" (Median score 6.0, p=<0.001) (**Table 14**). Six parameters scored a median score of 5 with statistically significant results (**Table 15**). These were "endoscopic view" (p=0.001), "polyp realism" (p=<0.001), "bleeding realism" (p=0.013), "haustral folds" (p=0.029), "anatomical realism of pedunculated polyps" (p=0.010), and "diathermy of the polyp" (p=0.026) (**Table 15**). Of the 15 parameters examined, only 3 were not statistically significant in favour of the simulator. These were, "anatomical realism of sessile polyps" (p=0.080), "resistance of scope movement" (p=0.406) and "paradoxical motion" (p=0.055).

[††] Full data set can be found in Appendix 2b

Clinical Level of participant	Number of participants
Consultant	8 (47.1%)
Post CCT	*2 (11.8%)*
Senior Trainee	6 (35.2%)
Nurse Specialist	1 (5.9%)
Type of procedure	**Mean no. of real life procedures (95% CI)**
Colonoscopies	371 (179-689)
Polypectomies	156 (35-355)
Biopsies	165 (42-360)
Type of simulation	**Mean no. of times simulator used (95% CI)**
Virtual Reality	64 (3-180)
Animal Model	4 (1-9)
Bench Model	8 (1-19)

Table 14: Demographic data from participants (CCT = completion of clinical training)

c) Overall score

The overall score for the simulation was statistically significant when compared to a neutral score (Median score 6.0, p <0.001). When participants were asked to compare the level of difficulty of the simulator compared to real life the result was not statistically significant (Median score 4, p=0.559). This indicates that a high degree of correlation between the simulator and real life experience.

Realism type	Realism parameter	Median score (IQR)	p-value
Visual	Mucosal Realism	6.0 (5.0–6.0)	0.001
	Endoscopic View	5.0 (5.0–6.0)	0.001
	Polyp realism	5.0 (5.0–6.0)	<0.001
	Bleeding realism	5.0 (4.0-6.0)	0.013
Anatomical	Haustral Folds	5.0 (4.0-5.0)	0.029
	Pedunculated polyps	5.0 (4.0-6.0)	0.010
	Sessile polyps	4.0 (4.0-5.0)	0.088
Mechanical	Resistance to scope	4.0 (4.0-5.0)	0.406
	Paradoxical motion	4.0 (3.0-6.0)	0.055
	Snare control	6.0 (5.0–6.0)	0.001
	Handling the polyp	6.0 (5.0–6.0)	0.001
	Diathermy of polyp	5.0 (4.0-6.0)	0.026
	Raising mucosa	6.0 (5.0-6.0)	<0.001
Summary	Overall simulation	6.0 (5.0-6.0)	<0.001
	Difficulty compared to reality	4.0 (3.0-4.0)	0.559

Table 15: Results from realism survey analysed using the Wilcoxon signed-rank for non-parametric data. Median values from the 7-point scale were compared with a hypothetical mean of 4 to determine statistical significance (significant parameters highlighted in grey)

4.4 Discussion

The results of this study have demonstrated that the WIMAT colonoscopy suitcase has a good level of content validity across a range of parameters. A cohort of participants, experienced in the skill of colonoscopy and polypectomy, awarded the model favourable scores for visual, anatomical and mechanical realism. All of the measured parameters for visual realism scored

well enough to produce a statistically significant result. Most encouragingly was the statistically significant score for the overall realism of the model and the non-statistically significant score when comparing the difficulty of the simulation to actual reality.

a) Non-statistically significant scoring

There were 3 parameters which did not reach statistical significance (**Table 15**). For anatomical realism, it was interesting to note that the scoring for simulated pedunculated polyps was more favourable than that for the simulated sessile polyps. However, when participants were asked to comment on the realism of performing a mucosal lift on a sessile polyp, a statistically significant favourable result was achieved. This would imply that the reduced level of anatomical realism of the sessile polyp did not significantly impact on the process of performing the polypectomy. The other non-statically significant parameters were for "resistance to scope movement" and "paradoxical motion". This may be a limitation of the current model. However, this should not significantly affect the use of the simulator which is designed to focus on polypectomy training as opposed navigation and endoscopic steering.

b) Conclusion

- The WIMAT colonoscopy suitcase demonstrates a good level of content validity
- Experts scored the model favourably in terms of visual, anatomical and mechanical realism using a previously validated survey
- Further work is needed to improve the anatomical realism of the simulated sessile polyps

Chapter 5: Validation

Testing the construct and concurrent validity of the WIMAT colonoscopy suitcase

5.1 Testing the construct and concurrent validity of the suitcase model: Introduction

The aim of this study is to evaluate the construct and concurrent validity of the WIMAT colonoscopy suitcase. *Construct validity is an agreement between a theoretical concept and a specific tool or procedure"* (Sedlack, 2011). A secondary aim is to test the hypothesis that the simulator can be reliably adjusted to re-create polypectomy task with varying degrees of difficulty.

5.2 Methods

a) Inclusion and exclusion criteria for participants

The Research Ethics Committee (REC) for Wales has stated that ethical approval was not required for this work under NHS research governance arrangements (**Appendix 3a**). The WIMAT Faculty and Trainee Database was used to identify and invite participants. All completed a questionnaire to establish current expertise and assign each subject to one of four groups (**Appendix 3b**). *Novices* had no prior experience of colonoscopy but were pursuing a career in either Gastroenterology or Surgery. *Intermediates* comprised of Surgical or Gastroenterology trainees, between the levels of Specialist Trainee 3-7 (ST3-7). These participants all had experience of performing <200 colonoscopies. This figure was chosen as the proposed number of colonoscopies needed to reach competency (JAG, 2011). *Advanced* participants were independent colonoscopists with full JAG accreditation. All routinely perform colonoscopy and polypectomy as part of their standard practice. *Experts* were highly skilled bowel screening colonoscopists. All work in teaching hospitals with high volume workloads. Previous endoscopic experience, simulator use, JAG courses attended or taught on, and real life DOPyS assessment scores were also recorded.

b) The simulation task

The set-up of the WIMAT colonoscopy suitcase was identical in each case, with 2 hypothesised degrees of difficulty as outlined in Chapter 3 of this thesis. All participants followed an identical study protocol. Candidates were asked to perform snare polypectomy and Roth net retrieval of polyp A and polyp B (**Figure 27**). Each candidate was allowed a maximum of 5 minutes to become familiar with the endoscopic equipment provided. The same endoscopic assistant was present during all procedures. All polypectomies were video recorded for future analysis and the files were coded in a random fashion using a number between 1 and 80.

c) Performance evaluation[‡‡]

Two JAG accredited colonoscopists, independent of the study design, analysed each performance using the DOPyS assessment form. Eight of 34 DOPyS assessment parameters were included according to their relevance to stalked polyps and generic skills (**Table 15**). DOPyS parameters were excluded if they could not be assessed by video format or were not relevant to stalked polypectomy. Performances were video recorded using the Archos 7 MN 6700 (Archos™, UK) from the video output of the endoscopic processor. Both assessors were given DOPyS descriptor guides for generic and stalked polyps (**Appendix 3c**). The time taken to complete each polypectomy was also recorded. Assessors remained blind to the level of experience of the participant. Upon completion of the evaluation stage, all candidates were awarded a DOPyS score (from 1-4) for Polyp A and Polyp B with their scores being compared between groups. Where available, "real-life" DOPyS scores were also collected from participant's so that they could be compared to their simulator DOPyS scores.

[‡‡] Following this task each participant was asked to complete a self-assessed DOPyS to allow comparison of these scores against experts ratings – these results are presented in Chapter 7 of this thesis

Generic	Score (0-4)
Optimises polyp position	
Optimising view of the polyp	
Directs snares accurately over polyp head	
Stalked Polyp	**Score (0-4)**
Places snare on appropriate position on the stalk	
Mobilises polyp to ensure appropriate amount of tissue is trapped within snare	
Applies appropriate degree of diathermy	
Post polypectomy	**Score (0-4)**
Retrieves, or attempts retrieval of polyp	
Overall competency at polypectomy	

Table 16: Adapted DOPyS video scoring system used for the video polypectomy assessment. Score: where 0= not done/unable to assess 1= Acceptable standards not yet met, frequent errors uncorrected 2 = some standards not yet met, aspects to be improved, some errors uncorrected 3= competent and safe throughout procedure, no uncorrected errors and 4 = highly skilled performance

d) Statistical Analysis

Previous research has demonstrated that experts (equivalent to our definition of experts) have an 88% (n=15) chance of scoring 3-4 (or pass) on the DOPyS, whereas non-experts (equivalent to our definition of intermediate level) have a 53% (n=8) chance of scoring between 3-4 (or pass) on the DOPyS (Gupta et al., 2011). Assuming that the novice group would take a similar drop in performance (from 53% to 18%), then 20 participants in each

group would give >80% power to detect a difference in performance of 35% between the groups, using a one-tailed test with a confidence level of 5%. A Wilcoxon signed-rank test was used to assess differences in group performance. Multiple group comparisons were assessed using Friedman's 2-way analysis of variance by ranks. Data is expressed as medians with an Inter Quartile Range (IQR) where a *P value* of <0.05 was considered statistically significant. Interrater reliability was compared by using the ***k*** statistic where 0.81-1.00 indicates very good agreement, 0.61-0.80 good agreement, 0.41-0.60 moderate agreement, 0.21-0.40 fair agreement and <0.20 poor agreement. All calculations were performed on PASW Statistics 18.[§§]

5.3 Results[***]

a) Demographic results

Eighty participants (20 per group) attempted the polypectomy task (**Table 16**). All novices had previously performed <50 colonoscopies and no pedunculated polypectomies. Sixteen (80%) intermediates had performed <100 colonoscopies and <50 pedunculated polypectomies (**Table 17**). In the advanced group, 15 (75%) participants had performed >500 colonoscopies and 14 (70%) >50 pedunculated polypectomies. All experts reported performing >500 colonoscopies and 14 (70%) >50 pedunculated polypectomies (**Appendix 3d**). The median number of times bench colonoscopy simulators were used by the cohort was 0.0 (1.0-3.0) (p=0.002 between groups) (**Table 18 / Appendix 3e**). For VR colonoscopy simulators this was 1.0 (0.0-3.0) (p=<0.001 between groups). The median number of times ex-vivo colonoscopy had been used previously, 0.0 (0.0-1.8) was not statistically significant between the groups (p=0.06). Seven (8.8%) participants (all novices) failed to complete the polyp A task and 7 (8.8%) (6 Novices, 1 Intermediate) failed to complete the polyp B task.

[§§] Statistical analysis has been reviewed and approved by Professor P Price
[***] Full dataset can be found in Appendix f-k

b) Simulation task completion times

Median (IQR) completion times (seconds) for Polyp A were 477 (322-672) for novices, 307 (215-395) for intermediates, 268 (235-379) for the advanced group and 331(199-451) for experts (p=0.058 between groups). For Polyp B, completion times were 683 (545-926) for novices, 550 (417-752) for intermediates, 434 (273-684) for the advanced group and 401 (298-524) for experts (p=<0.001 between groups) (**Figure 31**).

Demographics		**Group**				
		N (n=20)	**I (n=20)**	**A (n=20)**	**E (n=20)**	**Overall (n=80)**
Mean age (yrs) (Standard Error)		30.0 (1.28)	33.4 (0.85)	44.7 (1.30)	46.6 (1.33)	38.6 (0.99)
Gender (% Male)		85%	90%	80%	95%	87.5%
Speciality	Medicine	5%	30%	40%	85%	46%
	Surgery	95%	70%	60%	15%	54%

Table 17: Comparing age, gender and speciality between the groups (N = novices, I = intermediates, A = advanced and E = expert)

Demographics	Total	Group				
		N (n=20)	I (n=20)	A (n=20)	E (n=20)	Overall (n=80)
Previous number of colonoscopies	0-50	20	0	0	0	20
	51-100	0	16	0	0	16
	101-150	0	3	0	0	3
	151-200	0	1	0	0	1
	200-500	0	0	3	0	3
	500-1000	0	0	12	16	28
	>1000	0	0	5	4	9
Previous number of Flexible sigmoidoscopies	0-50	20	7	0	0	33
	51-100	0	6	0	0	15
	101-150	0	2	0	0	2
	151-200	0	2	1	0	3
	200-500	0	3	4	0	7
	500-1000	0	0	14	17	31
	>1000	0	0	1	3	4
Previous number of pedunculated polypectomies	0-50	20	20	6	6	52
	51-100	0	0	5	8	13
	101-150	0	0	4	1	5
	151-200	0	0	1	1	2
	200-500	0	0	0	0	0
	>500	0	0	4	4	8
Previous number of sessile polypectomies	0-50	20	20	6	6	52
	51-100	0	0	9	8	17
	101-150	0	0	0	1	1
	151-200	0	0	0	0	0
	200-500	0	0	4	0	4
	>500	0	0	1	5	6

Table 18: Previous number of "real-life" endoscopic procedures performed by each group

(N=novices, I=intermediates, A = advanced and E = expert)

		N (n=20)	I (n=20)	A (n=20)	E (n=20)	Overall
Total no. of times colonoscopy simulator used	VR	6	35	24	356	421
	EV	2	18	20	150	190
	CA	0	0	5	0	5
	LA	0	2	22	16	40
	BE	9	22	0	554	585
Total number of courses attended/taught	FN	0	10	10	10	30
	BA	0	12	14	16	42
	IN	0	0	4	13	17
	AD	0	1	9	16	26
	PY	0	0	6	13	19

Table 19: Previous experience of colonoscopy simulation and simulator courses attended/ taught on. N=novices, I=intermediates, A = advanced, E = expert, VR = virtual reality, EV = ex-vivo, LA = Live animal, BE = Bench model, FN = foundation skills, BA = basic skills in colonoscopy, IN = intermediate skills in colonoscopy, AD = advanced skills in colonoscopy PY = polypectomy training course

c) Comparing DOPyS scores for polyp A versus B

Overall median DOPyS scores (**Table 19**) for novices were 1.00 (1.00-1.87) for polyp A and 0.50 (0.00-1.00) for polyp B (p=<0.001). Intermediate group overall DOPyS scores were 2.50 (2.00-2.88) for polyp A and 2.00 (1.13-2.50) for polyp B (p=0.033). Advanced participants scored 3.00 (2.50-3.50) for polyp A and 2.50 (2.00-3.00) for polyp B (p=0.009). Expert

DOPyS scores were 3.00 (3.00-3.88) for polyp A and 3.00 (2.50-3.50) for polyp B (p=0.465). In the novice group, scores decreased from polyp A to polyp B in a statistically significant manner for all DOPyS parameters except for the "*directs snare accurately over polyp head*" variable (p=0.403). This was the same for the intermediate and advanced groups excluding the "*places the snare at appropriate position on the stalk*" variable (p=0.055 and p=0.331 respectively). There were no statistically significant difference between scores for polyp A and B for any parameter were exhibited in the expert group.

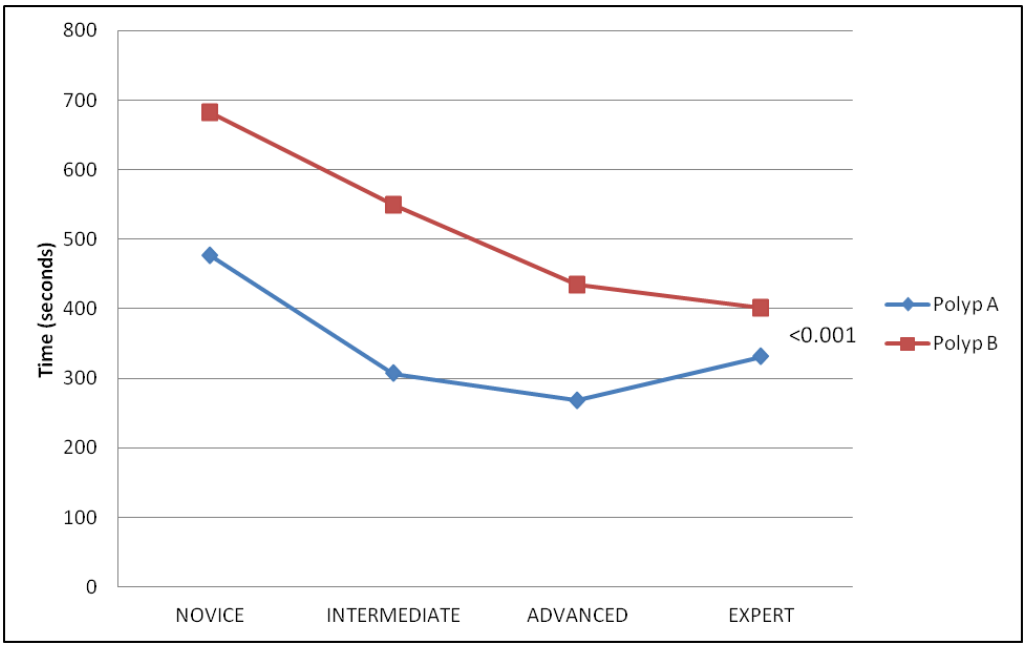


Figure 30: Median completion times for polyp A and polyp B across the groups, p-value generated from Friedman's analysis of variance

Group	Polyp task	DOPyS parameter assessed							
		Attempts to achieve optimal position	Optimises view by aspiration/ insufflation/ wash	Directs snare accurately over polyp head	Places the snare at appropriate position on the stalk	Appropriate amount of polyp tissue trapped	Applies appropriate degree of diathermy	Retrieves or attempts retrieval of polyp	Overall competency at polypectomy
Novices (n=20)	Polyp A	1.00 (1.00-1.87)	1.00 (1.00-1.50)	0.75 (0.00-1.50)	1.00 (1.00-1.50)	1.25 (1.00-1.50)	1.00 (1.00-1.50)	1.00 (0.50-1.50)	1.00 (1.00-1.87)
	Polyp B	1.00 (0.00-1.38)	1.00 (0.00-1.00)	1.00 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)
	P-value	<0.01	<0.01	0.40	<0.01	<0.01	<0.01	<0.01	<0.01
Intermediate (n=20)	Polyp A	2.25 (2.00-2.50)	2.25 (2.00-2.50)	2.00 (2.00-2.88)	2.50 (2.00-2.88)	2.50 (2.00-3.00)	2.50 (2.00-2.88)	2.50 (2.50-3.00)	2.50 (2.00-2.88)
	Polyp B	1.75 (1.50-2.50)	2.00 (1.50-2.38)	1.75 (1.50-2.50)	2.00 (1.13-2.5)	2.00 (1.50-2.50)	2.00 (1.13-2.50)	2.00 (1.00-2.50)	2.00 (1.13-2.50)
	P-value	<0.01	<0.01	0.01	0.06	0.01	0.04	<0.01	0.03
Advanced (n=20)	Polyp A	3.00 (2.50-3.50)	3.00 (2.13-3.50)	3.00 (2.63-3.50)	3.00 (2.00-3.50)	3.00 (2.50-3.50)	3.00 (2.50-3.50)	3.00 (2.13-3.50)	3.00 (2.50-3.50)
	Polyp B	2.50 (2.00-3.00)	2.50 (2.00-3.00)	2.50 (2.00-3.00)	2.75 (2.00-3.38)	2.50 (2.0-3.5)	2.75 (2.00-3.38)	2.25 (1.50-3.00)	2.50 (2.00-3.00)
	P-value	0.01	0.01	0.04	0.33	0.01	0.02	0.01	0.01
Experts (n=20)	Polyp A	3.25 (2.50-3.88)	3.00 (2.50-3.50)	3.00 (2.50-3.88)	3.00 (2.50-3.88)	3.00 (3.00-3.88)	3.00 (2.50-3.88)	3.25 (2.63-3.50)	3.00 (3.00-3.88)
	Polyp B	3.00 (2.50-3.50)	3.00 (2.50-3.38)	2.75 (2.50-3.50)	3.00 (2.50-3.50)	3.00 (2.50-3.50)	3.00 (2.13-3.50)	2.5 (2.50-3.50)	3.00 (2.50-3.50)
	P-value	0.07	0.46	0.31	0.69	0.18	0.64	0.16	0.47

Table 20: Median (IQR) DOPyS scores for Polyp A and B for 2 independent JAG accredited reviewers. Wilcoxon Signed-Rank test for non-parametric data used to assess differences between polyp A and B (p values were significant if <0.05, indicated by the grey boxes)

Group	Polyp task	DOPyS parameter assessed							
		Attempts to achieve optimal position	Optimises view by aspiration/ insufflation/ wash	Directs snare accurately over polyp head	Places the snare at appropriate position on the stalk	Appropriate amount of polyp tissue trapped	Applies appropriate degree of diathermy	Retrieves or attempts retrieval of polyp	Overall competency at polypectomy
N vs I	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
N vs A	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
N vs E	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
I vs A	Polyp A	<0.01	0.01	0.01	0.03	0.03	0.02	0.09	0.01
	Polyp B	0.01	<0.01	0.01	0.02	0.04	0.04	0.22	0.03
I vs E	Polyp A	<0.01	0.01	<0.01	0.01	<0.01	<0.01	0.03	<0.01
	Polyp B	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01
A vs E	Polyp A	0.29	0.91	0.86	0.50	0.50	0.74	0.95	0.46
	Polyp B	0.09	0.15	0.11	0.30	0.24	0.14	0.08	0.06

Table 21: Differences for DOPyS scores between groups for polyp A and B. Calculated using Mann-Whitney U (2 samples) test for non-parametric data (p values were significant if <0.05, indicated by the grey boxes). N=novices, I=intermediates, A=advanced, E=experts

d) Comparing DOPyS scores between the groups

Differences in DOPyS parameter scores between the groups were statistically significant in the majority of cases except for when the advanced group was compared with the experts (**Table 20**). Results were highly significant when novices were compared with either intermediates, advanced or experts for either polyp A or B. This was similarly the case when intermediate DOPyS scores were compared with the advanced or expert groups. Non-statistically significant parameters were, "*places the snare at appropriate position on the stalk*" (I vs A for polyp A, p=0.082) and "*retrieves or attempts retrieval of polyp*" (I vs A for polyp B, p= 0.239 and I vs E polyp A, p=0.114). No statistical difference was demonstrated for any parameter in either polypectomy between the advanced and expert groups (overall A vs E polyp A, p=0.711 and polyp B, p=0.107).

e) Concurrent validity (simulator versus "real-life" DOPyS scores)

Fifteen (75%) participants in the intermediate group submitted evidence of "real-life" DOPyS scores for level 1 pedunculated polypectomy. Median "real-life" DOPyS scores were comparable the median simulator scores (**Table 21**). Five out of eight parameters showed no statistical difference when comparing real vs simulator DOPyS scores ("*directs snare accurately over polyp head*" (p=0.125), "*places the snare at appropriate position on the stalk*" (p=0.085), *"mobilises polyp to ensure appropriate amount of tissue is trapped"* (p=0.063), *"applies appropriate degree of diathermy"* (p=0.072) and *"overall competency at polypectomy"* (p=0.069). DOPyS parameters that were statistically significant included, *"attempts to achieve optimal position"* (p=0.023) *"optimises view by aspiration/ insufflation/ wash"* (0.040) and "*retrieves or attempts retrieval of polyp*" (0.030).

Group	Polyp task	DOPyS parameter assessed							
		Attempts to achieve optimal position	Optimises view by aspiration/ insufflation/ wash	Directs snare accurately over polyp head	Places the snare at appropriate position on the stalk	Mobilises polyp to ensure appropriate amount of tissue is trapped	Applies appropriate degree of diathermy	Retrieves or attempts retrieval of polyp	Overall competency at polypectomy
Intermediates (I) Real Life DOPyS (n=15)	SIMULATOR DOPyS	1.75 (1.75-2.75)	1.75 (1.50-1.75)	2.00 (1.00-2.00)	2.00 (1.50-2.50)	2.00 (1.5-2.50)	2.00 (1.50-2.50)	2.00 (1.00-2.00)	2.00 (1.50-2.50)
	"REAL-LIFE" DOPyS	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.50-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	1.00 (1.00-2.00)	2.00 (1.75-2.50)	2.00 (1.00-2.00)
	P-value	0.02	0.04	0.13	0.09	0.06	0.07	0.03	0.07

Table 22: Comparison of *simulator DOPyS scores* versus *"real-life" DOPyS score* for pedunculated polypectomy using a Wilcoxon Signed-Rank test for non-parametric data, $p<0.05$ is significant. Statistically significant results are highlighted in grey

f) Interrater reliability

Analysis of Interrater reliability of the two JAG assessors showed moderate to very good agreement for overall DOPyS parameters for novices (polyp A ***k***=0.8 and polyp B ***k***=0.5). There was fair agreement for intermediate scores (polyp A ***k***=0.4 and polyp B ***k***=0.4) and advanced (polyp A ***k***=0.4 and polyp B ***k***=0.2) group scores. In the expert cohort this ranged from fair to moderate (polyp A ***k***=0.6 and polyp B ***k***=0.2). There was reduced agreement between assessors when scoring the more technically demanding polyp B[†††].

5.4 Discussion

a) Summary of results

This study demonstrates that the WIMAT colonoscopy suitcase has good construct and concurrent validity for skills training in colonoscopic polypectomy. DOPyS scores obtained on the simulator correlate with the "real-life" levels of expertise. Novice colonoscopists perform snare polypectomy on the simulator at an expected level of competency for their skills set. DOPyS scores improve on the simulator as the level of "real-life" experience increases. "Real-life" DOPyS scores also reflect the simulator scores obtained on the model.

b) Different degrees of difficulty for each polyp task

We hypothesised that placing polyps in differing positions inside the simulator would provide candidates with varying degrees of difficulty. The results indicate that this may be the case. The proposed, complex polyp B took longer for all groups to remove and generally resulted in lower DOPyS scores being obtained. No statistical difference was demonstrated between scores for polyp A and B in the expert group. However, expert scores were lower for polyp B coupled with an increase in the time taken to perform the task. Reduced inter-rater agreement

[†††] ***k*** = ***k*** statistic where 0.81-1.00 very good agreement, 0.61-0.80 good agreement, 0.41-0.60 moderate agreement, 0.21-0.40 fair agreement, <0.20 poor agreement

for polyp B DOPyS scores was also demonstrated. This correlated with increasing candidate "real-life" experience and may be because the DOPyS is not designed for assessing expert colonoscopists performing technically demanding polypectomy. Median scores across DOPyS parameters were similar for both polyp A and B. This may reflect that both assessors may have had a tendency to give similar scores for consecutive parameters when scoring individuals. Alternatively, it could provide further evidence for the construct of the model as wide variations between parameters would not be expected if the simulator is a true representation of real life. For example, if a candidate scores 1 for "*Optimising view of the polyp", "Directs snares accurately over polyp head" and "Places snare on appropriate position on the stalk"* a similar score should be expected for other parameters such as *"Mobilises polyp to ensure appropriate amount of tissue is trapped within snare"*. These results demonstrated that this was the case.

c) Inter-group assessment

Inter-group comparison demonstrated statistically significant differences across a range of DOPyS parameters. This was particularly apparent when novices were compared to all other groups and in the majority of parameters when intermediates were compared to either advanced or expert participants. Although no statistically significant difference was found when comparing advanced and expert groups, further refinements in polyp placement and complexity may allow this to be detected if a difference does in fact exist. The most consistently non-significant DOPyS parameter when comparing groups was for "*retrieves or attempts retrieval of polyp*". This may be explained by the way in which the simulated polyps are constructed. Polyps are made from a mix injected into porcine sausage skin forming the polyp stalk. If the snare is not placed around the stalk then the diathermy cuts through the mix and liquefies it. This causes the polyp to disintegrate, making retrieval difficult.

d) Concurrent validation

Comparison of simulator DOPyS scores with "real-life" scores showed no statistical difference for the majority of parameters including overall competency. This provides evidence for the concurrent validity of this model. Statistical differences in favour of "real-life" DOPyS scores were demonstrated for optimal positioning and optimising view. This could be a reflection of the model's ability to replicate real life scope handling or may be due to the difficulty in correlating retrospective "real-life" polypectomy tasks with the standardised simulated polypectomy being performed.

e) Limitations of the study

One limitation of this study is that it focuses on simple snare polypectomy. Although this allows reliable standardisation of the polypectomy tasks, it does not fully validate the model as useful for training other therapeutic colonoscopic procedures. Further work is required if the model is to be used as a training platform for more complex interventions such as Endoscopic Mucosal Resection (EMR) and Endoscopic Submucosal Dissection (ESD). This model has the benefit of being realistic, portable and cost-effective however, it is limited by the fact that it does not have inbuilt parameters for assessment. Therefore, retrospective DOPyS analysis of videoed performances was used. This restricted the number of DOPyS parameters that could be applied. If this model is to be fully utilised in training and assessment it may be more useful to establish quantitative forms of performance evaluation.

f) Conclusions

- WIMAT colonoscopy suitcase demonstrates construct and concurrent validity
- Endoscopists perform snare polypectomy on the simulator at an expected level of competency for their skills set
- Real-life DOPyS scores are comparable with simulator DOPyS scores
- The simulator can be adjusted to provide varying degrees of technical difficulty

Chapter 6: Skills Transfer

A prospective, randomised pilot study comparing ex-vivo with virtual reality simulation training in the transfer of skills to real-life colonic polypectomy

6.1 Introduction

a) Skills transfer in colonoscopy simulation training

One of the fundamental assumptions of simulation training is that the skills acquired in simulated environment should lead to improved performance in the clinical setting (Sturm et al., 2008). This concept is referred to as skills transfer. There have been 7 previous studies which focus on the evaluation of skills transfer in VR diagnostic lower GI endoscopy (**Table 22**). Several of these studies recruit low numbers of participants and use variable methodology including different training regimes and types of simulator. All studies showed that simulation-based training provides some advantage over no simulation-based training (Sturm et al., 2008). Three of these studies demonstrated that the benefit of simulation training is only apparent during the initial stages of learning, after which, no difference was found (Ahlberg et al., 2005, Cohen et al., 2006, Sedlack and Kolars, 2004). Endoscopic VR simulation can therefore be used to reduce the learning curve of a trainee but not replace patient-based teaching.

b) Skills transfer in therapeutic colonoscopy simulation training

There is no published evidence for skills transfer in ex-vivo colonoscopy simulation or in therapeutic lower GI endoscopy. This is particularly surprising as these procedures are technically challenging and have higher complication rates than diagnostic procedures (Dafnis et al., 2001). For therapeutic upper GI endoscopy, there are several papers which demonstrate skills transfer following simulation training. (Hochberger et al., 2005, Maiss et al., 2007). Haycock *et al* have recently compared knowledge-based teaching versus simulation training in 4 clinical scenarios; nonvariceal haemostasis, snare polypectomy and PEG tube insertion using the Erlangen Endo Trainer (ECE-Training GmbH, Germany) and

Oesophageal Stricture dilation using the upper GI phantom (Adam;Rouilly Ltd, UK) (Haycock et al., 2009b). This study concluded that simulator training was superior to knowledge-based training in terms of the level of skills transfer demonstrated.

c) How many hours of simulator training are needed to demonstrate skills transfer?

In previous skills transfer studies, the length of VR simulator colonoscopy training ranges from 6 to 20 hours (Sturm et al., 2008). Sedlack *et al* concluded that 3 hours of simulator training was insufficient to demonstrate skills transfer in diagnostic colonoscopy (Sedlack et al., 2004). Other research showed that skills transfer is exhibited following 10 hours of simulator training (Tuggy, 1998). In contrast, there are currently no papers which detail the minimum simulator training time needed to demonstrate skills transfer in colonic polypectomy.

d) How many simulator cases are needed to demonstrate skills transfer?

The minimum number of "real-life" colonoscopy cases needed to reach proficiency can be as high as 200 (Cass et al., 1993). However, there is no definitive evidence which states the minimum number of simulator cases needed to maximise the potential of colonoscopy training. Several studies demonstrate a significant difference in patient discomfort scores during the first 15 VR colonoscopies performed (Ahlberg et al., 2005, Sedlack and Kolars, 2004, Sedlack et al., 2004). However, Gerson *et al* showed no difference in levels of patient discomfort between simulator trainees and non-simulator trainees independent of number of simulator cases performed (Gerson and Van Dam, 2003). These papers are however, difficult to compare due the lack of standardisation in the levels of sedation and pain relief administered.

Author	Simulator	Comparison	n	Simulator training	Outcome measures	Findings
(Ahlberg et al., 2005)	Accutouch	VR Simulation vs no simulation	12	Until pre-defined expert level of performance reached	Patient discomfort scores for first 10 colonoscopies	Favours simulation training
(Sedlack and Kolars, 2004)	Accutouch	VR Simulation vs no simulation	8	6 hour simulator program	Insertion, completion,% mucosa visualised, patient discomfort for first 15 colonoscopies	Favours simulation up to first 30 live cases
(Sedlack et al., 2004)	Accutouch	VR Simulation vs no simulation	38	3 hour simulator program	Trainers rated performance and patient completed pain scores	No difference in training rating and pain scores favoured simulation group
(Cohen et al., 2006)	Simbionix	VR Simulation vs no simulation	49	10 hours simulator program (unsupervised)	Competency scores and patient comfort for first 200 colonoscopies	Favours simulation training in early stages of training
(Tuggy, 1998)	Gastro-sim	VR Simulation vs no simulation	10	6 hour simulator program	Insertion time, completion time, %mucosa visualised and viewing quality	Favours simulation training
(Gerson and Van Dam, 2003)	VR simulator	Simulation vs patient based	16	Undefined simulator program	Duration, completion, ability to perform retro flexion, and level of patient comfort/discomfort for first 5 sigmoidoscopies	Favoured patient teaching apart from time and comfort scores where there was no difference
(Haycock et al., 2010)	Olympus	Simulation vs patient based	36	16 hour simulator program	Simulator metrics, blinded DOPS and Global Scores, 3 live cases before and after	Simulation group matched patient training

Table 23: Summary of previous skills transfer studies in colonoscopy/sigmoidoscopy simulation. Adapted from (Sturm et al., 2008)

e) Aims of this trial

There is a lack of evidence in the literature to assess the benefit of skills transfer from ex-vivo polypectomy simulation training. There are also no papers which demonstrate the number of simulator cases or training hours needed. The aim of this study was to conduct a small-scale RCT pilot trial to assess the feasibility of comparing the skills transfer obtained from ex-vivo versus VR polypectomy simulation training.

6.2 Methods and Materials

a) Clinical Trials database registration, approval and design

This pilot has been designed in accordance with published guidelines on conducting pilot work by Thabane *et al* (Thabane et al., 2010). The study was registered on the International Standard Randomised Controlled Trial Number database (www.isrctn.org) on the 26/10/2011, and allocated the reference identifier ISRCTN41736707. The study protocol was approved by the School of Postgraduate Medical & Dental Education Research Ethics Committee, Cardiff University (**Appendix 4a / 4b**). A single centre, randomised, double-blind pilot study was conducted at the Welsh Institute for Minimal Access Therapy (WIMAT) between September 2012 and July 2013.

b) Participant selection

Participants were recruited from the Wales Deanery via electronic invitation. Potentially suitable participants were identified using the Wales Deanery National Database for Specialist Training according to the following inclusion and exclusion criteria.

c) Inclusion and exclusion criteria

In order to perform "real-life" polypectomy, trainees require a certain level of expertise in endoscopic handling. It was therefore, decided that eligible participants should be those already enrolled in specialty training (ST 3-7) with a designated speciality training number in either General Surgery or Gastroenterology. Experience of performing either flexible sigmoidoscopy or colonoscopy was mandatory, but candidates were not included if they had been provisionally or fully accredited to perform either of these skills independently. In order to quantify these criteria, only trainees who had performed <200 colonoscopies and <50 pedunculated polypectomies were included. Participants were excluded if they had undertaken formal teaching on either simulator previously (**Table 23**). Consultant trainers were notified once a trainee was recruited (**Appendix 4c**).

d) Assessing eligibility

Candidates who responded to the initial electronic invitation were asked to complete a baseline electronic questionnaire (**Appendix 4d**) to confirm eligibility. Participant demographics, previous endoscopic experience and simulator training courses attended were recorded. Where available, participants were asked to submit the results of their Direct Observation of Polypectomy Skills (DOPyS) scores from previous "real-life" polypectomies performed. This was intended to allow an average baseline DOPyS score to be calculated alongside previous endoscopic and simulator experience. This ensured that there was a reasonable level of homogeneity in the cohort before randomisation.

e) Consent

All participants who met the inclusion criteria were invited to the WIMAT centre to complete the trial induction process. Written informed consent was obtained from each participant

before recruitment into the study. The Informed Consent Form (ICF), (**Appendix 4e**) approved by the School of Postgraduate Medical & Dental Education Research Ethics Committee was used throughout the study. A duplicate of the signed ICF was provided to each study participant.

Inclusion criteria	Exclusion criteria
National Training Number (NTN) in either General Surgery or Gastroenterology	Provisionally or fully accredited to perform flexible sigmoidoscopy or colonoscopy
Previous experience of performing <200 colonoscopies and <50 pedunculated polypectomies	Previous experience of formal teaching on either the WIMAT colonoscopy suitcase or the Simbionix GI mentor II simulators
Successful completion of a JAG Basic Endoscopy Course	Unable to commit to simulator training program; three sessions

Table 24: Summary of the inclusion and exclusion criteria

f) Randomisation process

Immediately after enrolment, each candidate was assigned a study identification (ID) number, chronologically, from a specified series (01-20). At this point trainees were block randomised using an electronic platform (http://www.sealedenvelope.com) on a 1:1 ratio into either Group A or Group B. The participant ID and group were automatically emailed to the chief investigator who recorded these details for future analysis (**Table 24)**

g) Simulator intervention

All participants were required to complete a total of 12 simulated polypectomy procedures over a 4 week period. This was divided into three separate self-directed training sessions,

with 4 polypectomy tasks per session. Candidates that were unable to complete all training sessions within 4 weeks were excluded from the analysis. Each session was a maximum of 1 hour in duration. Participants in Group A performed all procedures using the GI mentor II VR colonoscopy simulator (Simbionix, USA) whereas participants in Group B used the WIMAT colonoscopy suitcase. **Figure 32** summarises this process.

Participant ID	Group
01	A
02	B
03	B
04	B
05	A
06	A
07	B
08	B
09	B
10	B
11	A
12	A
13	A
14	A
15	A
16	B
17	B
18	A
19	A
20	B

Table 25: Results from the randomisation process. Group A represents VR simulation training and Group B represents Ex-vivo simulation training

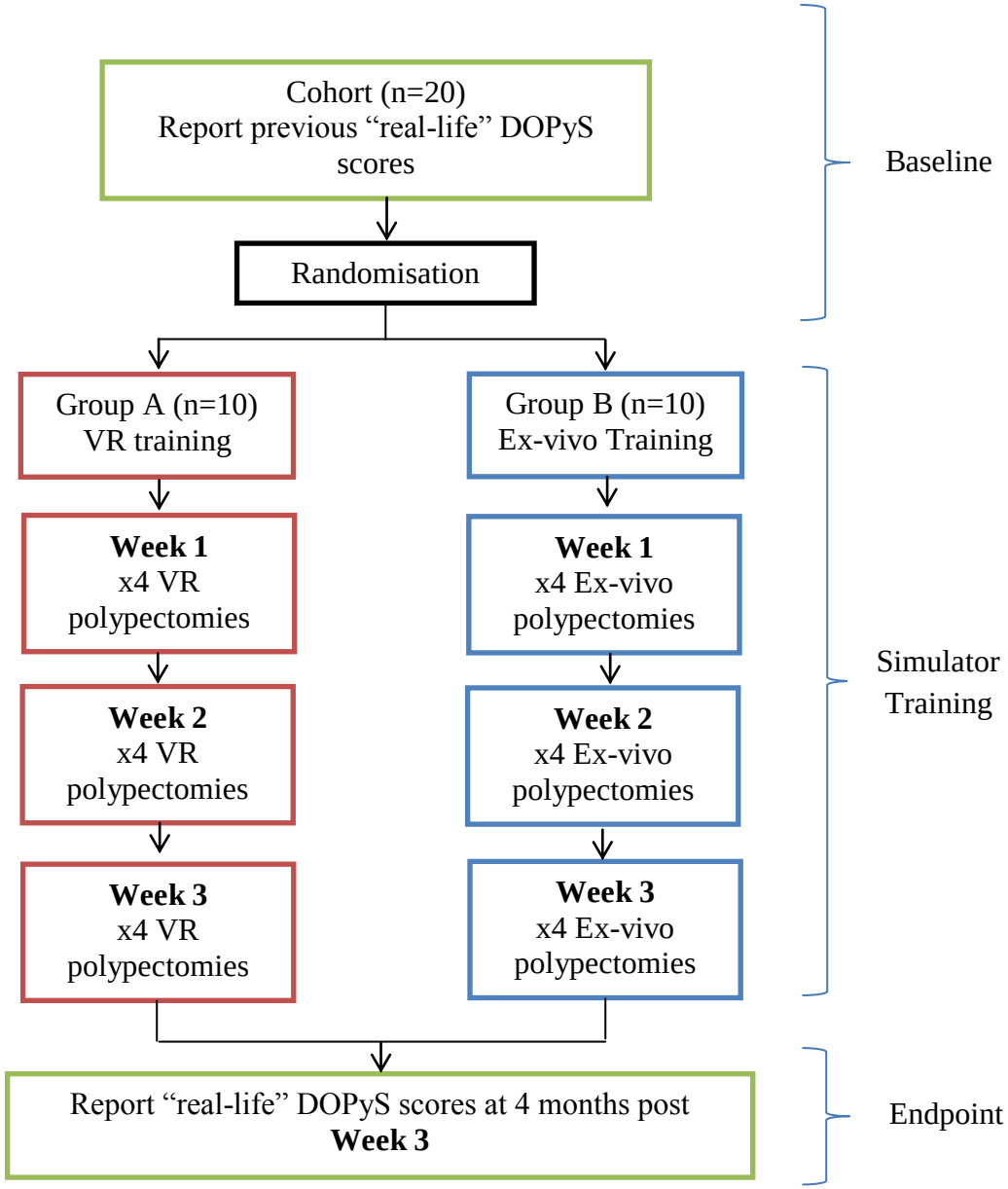


Figure 31: Summary of study design

h) Group A polypectomy task

Group A used the GI Mentor II (Simbionix, USA) to complete 4 previously validated VR pedunculated polypectomy tasks in the same order during each training session; Module I *Case 5* (1.5), Module I *Case 6* (1.6), Module I *Case 7* (1.7) and Module II *Case 1* (2.1). The first polypectomy (1.5) performed during each session was video recorded for analysis (**Table 25**).

Module	Description
1.5	Long sigmoid with tumour. Pedunculated polyp in descending colon. Long and twisted hepatic flexure
1.6	Long sigmoid with ischaemic colitis. Descending colon with diverticulum. Splenic flexure with pedunculated polyp. Very redundant transverse colon. Caecum which is shifted to mid abdomen
1.7	Long sigmoid. Long and twisted hepatic flexure. Sessile polyp in the transverse colon. Pedunculated polyp in the descending colon
2.1	Pedunculated polyp in the sigmoid colon, 12mm in diameter

Table 26: Description of the VR tasks performed by Group A on the GI mentor II during each training session (Simbionix, 2012)

i) Group B polypectomy task

Group B completed 4 standardised polypectomies on level 1 (0-1p) pedunculated polyps. All polyps were constructed according to the following dimension; diameter = 10mm, length = 15mm, stalk = 5mm. The positions of polyps A and B from the previously validated task outlined in chapter 3 of this thesis were used. In addition 2 further polyps (C and D) were inserted into the porcine bowel (**Figure 33** and **Table 26**)

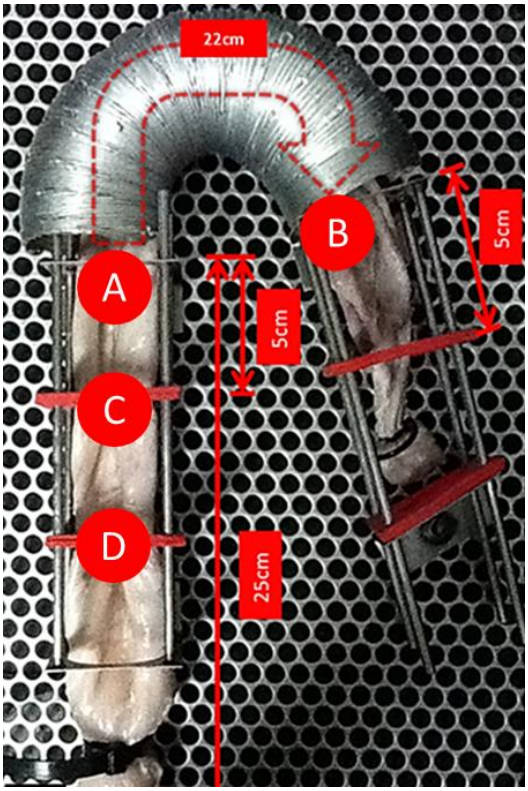


Figure 32: WIMAT colonoscopy suitcase internal set-up for skills transfer trial

Module	Description
A	10mm diameter polyp placed at 25cm from anal verge in a 6 o' clock position, proximal to a luminal fold on the up-going part of bowel
B	10mm diameter polyp placed at 47cm from anal verge in a 1 o' clock position, behind a luminal fold, distal to a colonic bend
C	10mm diameter polyp placed at 20cm from anal verge in a 1 o' clock position, behind a luminal fold, distal to a colonic bend
D	10mm diameter polyp placed at 15cm from anal verge in a 1 o' clock position, behind a luminal fold, distal to a colonic bend

Table 27: Description of the Ex-vivo tasks performed by Group B on the WIMAT colonoscopy suitcase during each training session

j) Assessment process

The first polypectomy performed during each of the 3 training sessions was recorded. Each video was then edited to demonstrate the polypectomy procedure. **Table 27** demonstrates the order of the simulated tasks undertaken and the polypectomy that was video recorded for analysis. Each video was then labelled with participant's ID (**Figure 34**).

	Group A				Group B			
	Simbionix Modules undertaken				Ex-vivo polyp Task undertaken			
Task	1.5	1.6	1.7	2.1	A	B	C	D
Order	1st	2nd	3rd	4th	2nd	1st	3rd	4th
Video Recorded	✓	-	-	-	-	✓	-	-

Table 28: Schematic of the format of polypectomy simulator training sessions

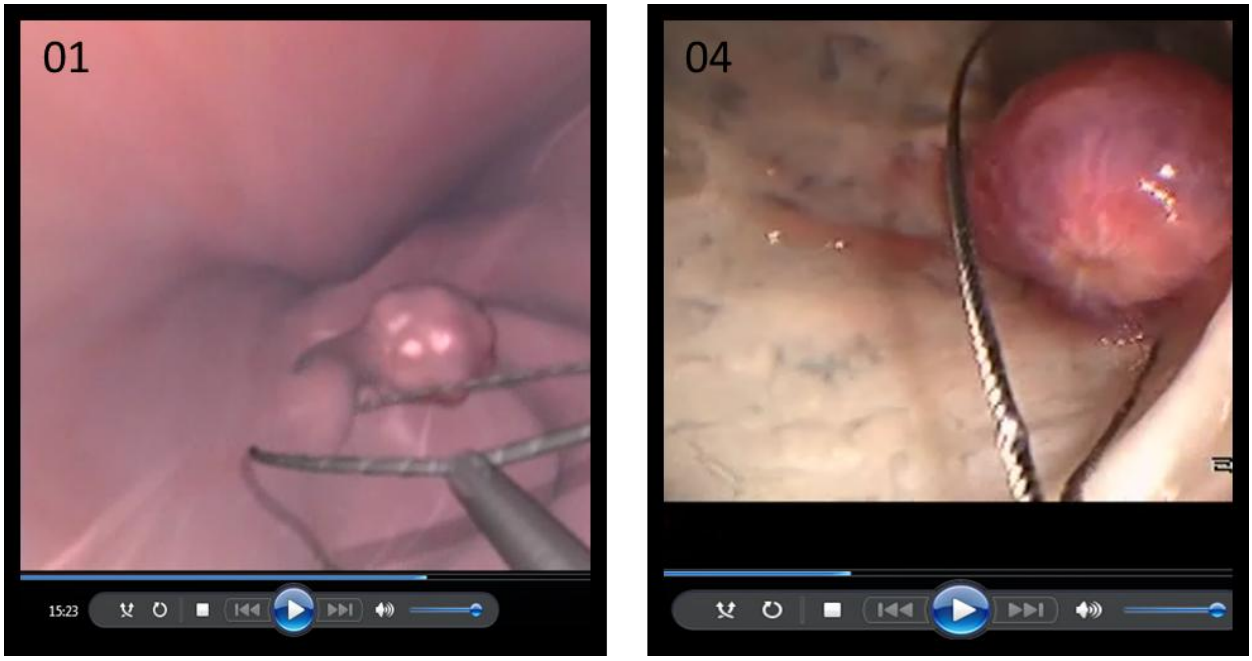


Figure 33: Screenshot of the assessors view for Group A (left) and Group B (right)

Two independent assessors (A and B), blind to participant identity scored the videos. Both assessors were experienced colonoscopists, accredited by JAG. Scoring was conducted using the shortened DOPyS assessment used during the construct validation trial (**Table 16**).

k) Follow up

All participants were scheduled for a 4 month follow-up after completion of week 3 of training. An electronic questionnaire was used to record any endpoint "real-life" DOPyS scores obtained during this time period (**Appendix 4f**). Variations in training including the number of procedures performed and additional training courses attended were also recorded for both groups.

l) Outcomes

In accordance with the recommendations for conducting a pilot RCT the following primary, secondary and feasibility outcomes were generated (Thabane et al., 2010). The primary outcome measure was the difference between baseline "real-life" DOPyS scores and endpoint "real-life" DOPyS scores. The secondary outcomes were:

1. The difference between endpoint "real-life" DOPyS scores for Group A and Group B
2. The difference between simulator DOPyS scores for Group A and Group B
3. The inter-rater correlation between Assessor A and Assessor B
4. The difference in simulated polypectomy procedure time for Group A and Group B

In addition, the following feasibility outcomes were established:

- The feasibility of recruitment and the suitability of inclusion and exclusion criteria
- Participant compliance with simulation training regimes
- The amount of simulation training required to demonstrate skills transfer
- The feasibility of primary and secondary endpoints

m) Statistical analysis plan

As this is a pilot study, a formal sample size calculation was not performed (Thabane et al., 2010). There are no studies which compare the use of two different endoscopic simulators with skills transfer as the primary endpoint. There are also no skills transfer studies that focus solely on polypectomy. Most evidence in the simulator literature compares VR platforms versus no extra training or patient based training as control groups. In these trials participant numbers range from 8 to 49. The study most closely related study to our own pilot was Haycock *et al*, who compared diagnostic colonoscopy simulation versus patient based

assessment using endpoint DOPS scores to evaluate skills transfer. In this full RCT study, 32 participants (16 per group) were enrolled. This study was powered to detect a 5-unit (1 standard deviation) difference between the 2 groups with a 5% significance level and 80% power (Haycock et al., 2010). For our pilot study it was therefore, estimated that a cohort of 20 participants (10 per group) would be sufficient in order to assess primary, secondary and feasibility outcomes. Summary statistics are presented as medians values with an interquartile range. This study was not sufficiently powered to perform meaningful statistical comparison testing. All statistical analysis was performed using SPSS version 20 (SPSS Inc, Chicago, IL).

6.3 Results

a) Demographics and disposition of participants[‡‡‡]

Figure 35 shows the number of participants randomised into the study and the number who were withdrawn and analysed. Twenty participants completed the baseline questionnaire and were randomised into either VR or ex-vivo training. In Group A (VR) two participants (ID 06 and 12) were unable to complete 3 sessions of training within the allotted time and were therefore, excluded. In Group B (ex-vivo) one participant (ID 03) was unable to complete 3 sessions of training within the allotted time and was therefore, excluded. During the follow-up phase one candidate from Group A (VR) (ID 11) and one from Group B (ID 17) did not respond to the electronic data collection tool within the allotted time period and were therefore excluded. An attempt was made to recruit replacement candidates for those excluded on 3 separate occasions, however, this was unsuccessful. A total of 7 candidates from Group A and 8 from Group B responded to the 4-month follow-up questionnaire and were therefore, included in the analysis. Only 2/7 candidates in Group A and 4/8 candidates in Group B had completed endpoint "real-life" DOPyS scores. These candidates were subject

[‡‡‡] Full baseline demographic results can be found in Appendix 4g-4j

to a further subgroup analysis.

b) Baseline summaries

Baseline demographics between the two groups were analogous (**Table 28**). Group A and Group B reported equivalent baseline lower GI endoscopic experience. Previous endoscopic simulator experience was also similar. Baseline DOPyS scores were identical in both groups.

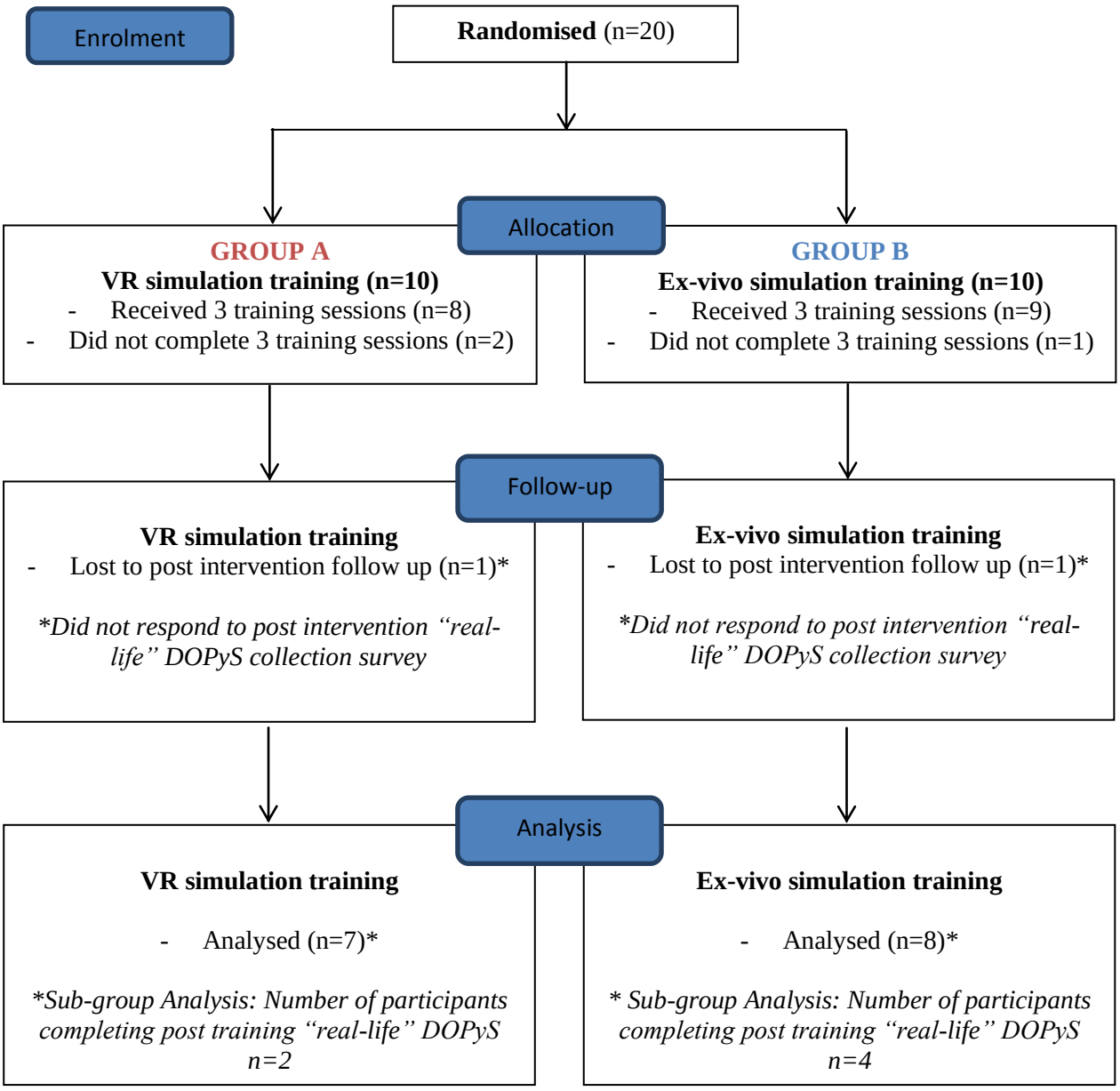


Figure 34: Study CONSORT diagram

Demographics	Group A (VR) (n=7)	Group B (Ex-vivo) (n=8)
Age (mean, SE)	33.38 (1.45)	33.50 (1.15)
Previous number of flexible sigmoidoscopies performed (median IQR)	50.00 (27.00-92.50)	50.00 (23.50-127.00)
Previous number of colonoscopies performed (median IQR)	70.00 (9.75-95.00)	58.00 (10.50-117.50)
Previous number of OGD's performed (median IQR)	210.00 (26.25-248.50)	100.00 (17.50-350.00)
Previous number of sessile polypectomies performed (median IQR)	2.00 (0.00-17.75)	0.00 (0.00-14.00)
Previous number of pedunculated polypectomies performed (median IQR)	3.50 (0.25-14.75)	7.00 (0.00-10.00)
Number of times bench endoscopic simulator used (median IQR)	0.50 (0.00-1.00)	1.00 (0.00-1.00)
Number of times VR endoscopic simulator used (median IQR)	0.00 (0.00-0.75)	0.00 (0.00-1.00)
Number of times ex-vivo endoscopic simulator used (median IQR)	0.00 (0.00-0.00)	0.00 (0.00-1.00)
Baseline overall competency at polypectomy DOPyS score (median IQR)	0.00 (0.00-0.75)	0.00 (0.00-0.75)

Table 29: Participant baseline demographics for Groups A and B

c) Primary outcome

i. Difference between baseline and endpoint "real-life" DOPyS scores

"Real-life" endpoint DOPyS scores showed some improvement at 4-months post simulator training for both groups. However, our analysis is restricted by the fact that not all candidates completed an endpoint "real-life" DOPyS assessment. Although electronic correspondence was returned by 7 candidates in Group A and 8 candidates in Group B only 2/7 in Group A and 4/8 in Group B submitted "real-life" polypectomy scores. Therefore, in the full analysis the majority of participants scored 0 (**Table 29**). A subgroup analysis was performed to evaluate the effect of simulator training on those who were able to submit endpoint "real-life" DOPyS scores. In Group A, these were candidates ID13 and ID15 and in Group B candidates ID 4,7,10 and 16. In some cases, several "real-life" DOPyS were submitted by individuals (**Table 30**). In all 6 participants there was an improvement in their DOPyS scores (**Table 31** and **Figure 36**).

Of the 18 "real-life" endpoint DOPyS scores submitted, 9 (50%) were from the sigmoid colon, 2 (11%) from the descending colon, 2 (11%) from the transverse 2 (11%) from the ascending colon and 1 (6%) from the rectum. For 2 (11%) polypectomies, the site was not specified. The overall mean size of the polyps resected was 7.80mm (SE 0.86) (**Table 31**).

DOPyS parameters	Baseline DOPyS (median IQR)	Endpoint DOPyS (median IQR)
Attempts to achieve optimal polyp position	0.00 (0.00-0.75)	0.00 (0.00-2.25)
Optimises view by aspiration / insufflation / wash	0.00 (0.00-0.75)	0.00 (0.00-2.25)
Determines the full extent of the lesion	0.00 (0.00-0.75)	0.00 (0.00-1.50)
Uses appropriate polypectomy technique	0.00 (0.00-1.00)	0.00 (0.00-2.25)
Adjusts / stabilises scope position	0.00 (0.00-1.00)	0.00 (0.00-2.00)
Checks all polypectomy equipment available	0.00 (0.00-1.00)	0.00 (0.00-1.75)
Checks snare closure prior to introduction into the scope	0.00 (0.00-1.00)	0.00 (0.00-1.50)
Clear instructions to and utilisation of endoscopy staff	0.00 (0.00-1.00)	0.00 (0.00-2.25)
Checks diathermy settings are appropriate	0.00 (0.00-0.50)	0.00 (0.00-2.00)
Photo-documents pre and post polypectomy	0.00 (0.00-1.00)	0.00 (0.00-2.25)
Applies prophylactic haemostatic measures if appropriate	0.00 (0.00-1.00)	0.00 (0.00-0.50)
Selects appropriate sized snare	0.00 (0.00-1.00)	0.00 (0.00-2.00)
Directs snare accurately over polyp head	0.00 (0.00-1.00)	0.00 (0.00-2.00)
Correctly selects en-bloc/ piecemeal removal depending on size	0.00 (0.00-1.00)	0.00 (0.00-1.00)
Advances snare sheath towards stalk as snare closed	0.00 (0.00-1.00)	0.00 (0.00-2.25)
Places snare at appropriate position on	0.00 (0.00-1.00)	0.00 (0.00-1.75)
Mobilises polyp to ensure appropriate amount of tissue trapped	0.00 (0.00-1.75)	0.00 (0.00-2.25)
Applies appropriate degree of diathermy	0.00 (0.00-1.00)	0.00 (0.00-2.25)
Examines remnant stalk/ polyp base	0.00 (0.00-0.00)	0.00 (0.00-2.25)
Identifies and appropriately treats residual polyp	0.00 (0.00-0.00)	0.00 (0.00-0.50)
Identifies bleeding and performs endoscopic haemostasis	0.00 (0.00-0.00)	0.00 (0.00-0.00)
Retrieves, or attempts retrieval of polyp	0.00 (0.00-0.00)	0.00 (0.00-2.25)
Checks for polyp retrieval	0.00 (0.00-0.00)	0.00 (0.00-2.25)
.Places tattoo completely, where appropriate	0.00 (0.00-0.00)	0.00 (0.00-0.50)
Overall competency at Polypectomy	0.00 (0.00-0.00)	0.00 (0.00-1.50)

Table 30: Comparison of “real-life” baseline and endpoint DOPyS scores for total cohort

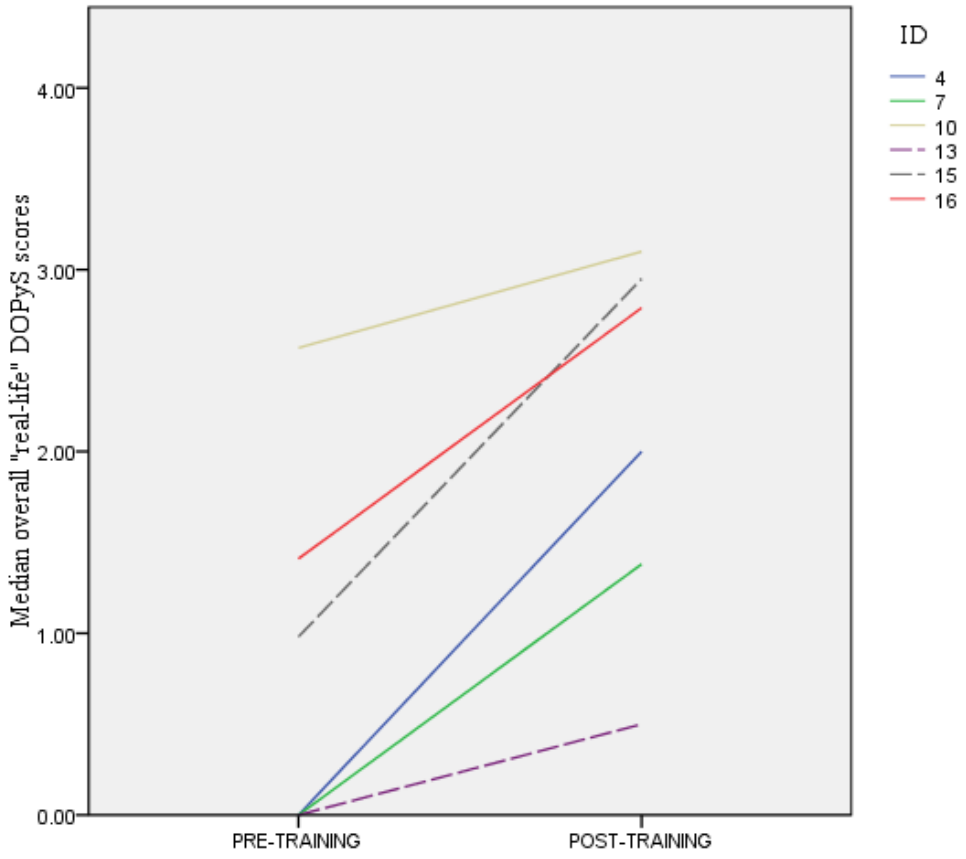


Figure 35: Line chart illustrating the median (IQR) baseline and endpoint scores for participants who completed the endpoint "real-life" DOPyS assessment (ID key represents individual participants)

ID	Baseline DOPyS (median IQR)	Endpoint DOPyS (median IQR)
4	0.00 (0.00-0.00)	2.00 (1.50-2.25)
7	0.00 (0.00-0.00)	1.50 (1.50-1.75)
10	3.00 (2.25-3.00)	3.00 (3.00-3.25)
13	0.00 (0.00-0.00)	0.50 (0.50-0.50)
15	1.00 (1.00-1.00)	3.00 (3.00-3.63)
16	1.50 (1.50-1.50)	3.25 (2.88-3.50)

Table 31: Comparison of the median (IQR) baseline and endpoint DOPyS scores for participants who completed the endpoint "real-life" DOPyS assessment

Group	ID	No. of polypectomies	Polyp size (mm)	Polyp site
A	13	1	9	Sigmoid Colon
A	15	4	8	Descending Colon
			6	Transverse Colon
			15	Sigmoid Colon
			9	Ascending Colon
B	4	3	9	Sigmoid Colon
			5	Sigmoid Colon
			15	Sigmoid Colon
B	16	4	12	Sigmoid Colon
			6	Descending Colon
			6	Ascending Colon
			12	Sigmoid Colon
B	7	2	5	Not recorded
			2	Not recorded
B	10	4	7	Rectum
			5	Sigmoid Colon
			4	Sigmoid Colon
			6	Transverse Colon

Table 32: Details of the nature of polypectomies performed at 4 months post simulator training for the total cohort

d) Secondary outcomes

i. The difference between endpoint "real-life" DOPyS scores for Group A and Group B

In general, median endpoint "real-life" DOPyS were higher across all assessment parameters in Group B compared with Group A (**Table 32**). However, these scores are heavily influenced by the poor level of submission (2 participants in Group A and 4 participants in Group B) of endpoint "real-life" DOPyS scores. However, when comparing the median endpoint "real-life" DOPyS of those candidates who did complete the endpoint assessment, the trend remains the same (**Table 33** and **Figure 37**). Although Group B baseline scores were lower than Group A, there endpoint scores were considerable higher. As the number of endpoint "real-life" DOPyS scores was low, this outcome needs to be interpreted accordingly. However, in a study with larger numbers of participants it would be interesting to evaluate if ex-vivo simulation is superior in terms of polypectomy skills transfer over VR models.

Overall competency DOPyS Score	**Group A / VR (n=2)**	**Group B / Ex-vivo (n=4)**
Baseline "real-life" DOPyS score	0.25 (0.00-1.00)	0.13 (0.00-1.50)
Endpoint "real-life" DOPyS score	0.50 (0.50-3.00)	2.25 (1.50-3.25)

Table 33: Comparison of the median (IQR) baseline and endpoint "real-life" DOPyS scores for Group A and Group B for participants who completed the endpoint "real-life" DOPyS assessment

Endpoint "real-life" DOPyS parameters	Group A / VR (n=7)	Group B / Ex-vivo (n=8)
Attempts to achieve optimal polyp position	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Optimises view by aspiration / insufflation / wash	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Determines the full extent of the lesion	0.00 (0.00-0.50)	1.5 (1.50-3.00)
Uses appropriate polypectomy technique	0.00 (0.00-0.50)	1.5 (1.50-3.00)
Adjusts / stabilises scope position	0.00 (0.00-0.50)	1.75 (0.00-3.00)
Checks all polypectomy equipment available	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Checks snare closure prior to introduction into the scope	0.00 (0.00-0.50)	0.75 (0.00-1.75)
Clear instructions to and utilisation of endoscopy staff	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Checks diathermy settings are appropriate	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Photo-documents pre and post polypectomy	0.00 (0.00-0.50)	1.50 (0.00-3.25)
Applies prophylactic haemostatic measures if appropriate	0.00 (0.00-0.00)	0.00 (0.00-1.50)
Selects appropriate sized snare	0.00 (0.00-0.50)	1.75 (0.00-3.00)
Directs snare accurately over polyp head	0.00 (0.00-0.50)	1.75 (0.00-3.00)
Correctly selects en-bloc/ piecemeal removal depending on size	0.00 (0.00-0.00)	1.00 (0.00-3.25)
Advances snare sheath towards stalk as snare closed	0.00 (0.00-0.50)	1.50 (0.00-3.25)
Places snare at appropriate position on	0.00 (0.00-0.50)	1.50 (0.00-3.00)
Mobilises polyp to ensure appropriate amount of tissue trapped	0.00 (0.00-0.50)	1.50 (0.00-3.25)
Applies appropriate degree of diathermy	0.00 (0.00-0.50)	1.75 (0.00-3.25)
Examines remnant stalk/ polyp base	0.00 (0.00-0.50)	0.00 (0.00-3.25)
Identifies and appropriately treats residual polyp	0.00 (0.00-0.00)	0.00 (0.00-1.75)
Identifies bleeding and performs endoscopic haemostasis	0.00 (0.00-0.00)	0.00 (0.00-1.25)
Retrieves, or attempts retrieval of polyp	0.00 (0.00-0.50)	1.75 (0.00-3.00)
Checks for polyp retrieval	0.00 (0.00-0.50)	1.75 (0.00-3.00)
.Places tattoo completely, where appropriate	0.00 (0.00-0.50)	0.00 (0.00-1.00)
Overall competency at Polypectomy	0.00 (0.00-0.50)	1.25 (0.00-2.75)

Table 34: Comparison of the median (IQR) endpoint "real-life" DOPyS scores by Group

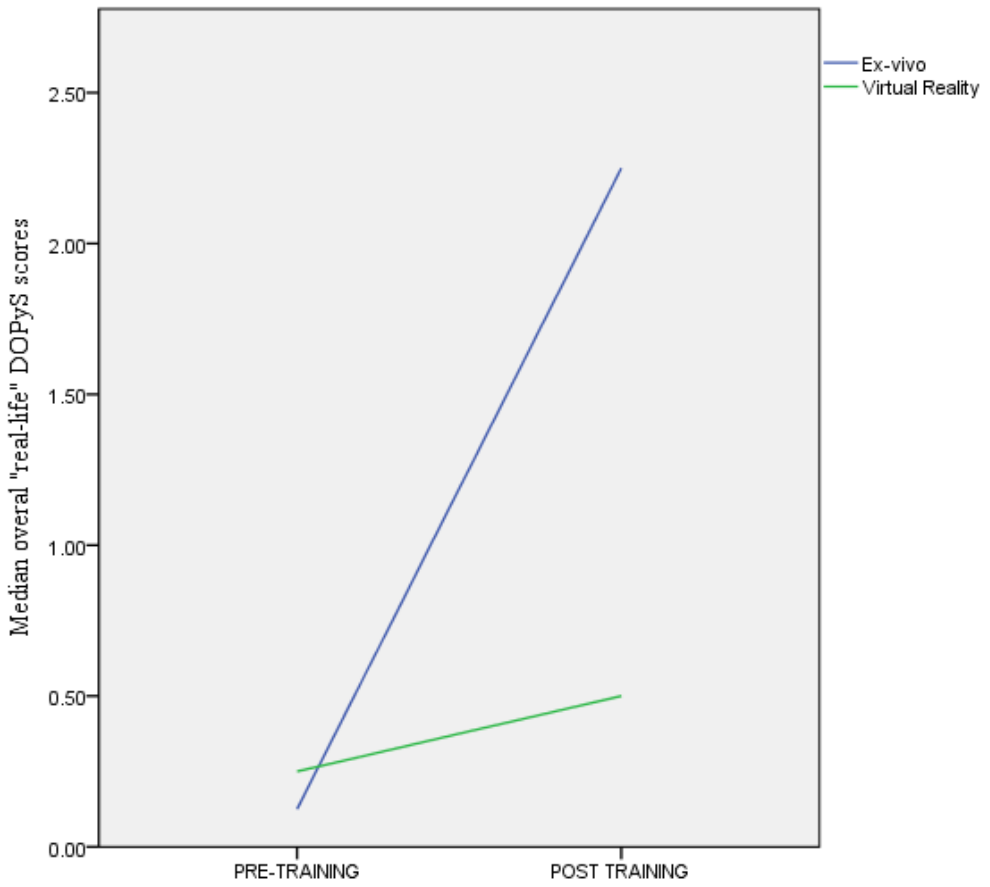


Figure 36: Comparison of the median (IQR) baseline (pre-training) and endpoint (post-training) "real-life" DOPyS scores for Group A / VR and Group B / Ex-vivo for participants who completed the endpoint "real-life" DOPyS assessment. This shows a greater level of improvement in DOPyS score following training using the Ex-vivo simulator

Potential co-founding variables which may have influenced the difference in endpoint DOPyS scores were; the types of "real-life" polypectomy performed and the clinical experience of each candidate during 4 month follow-up period. **Table 34** shows similar size and location of "real-life" polypectomies were performed between the groups. **Table 35** demonstrates that Group A in general, did attend more colonoscopy training lists and performed a greater number of colonoscopic procedures. These co-founding variables will need to be closely controlled in a larger trial.

ID	Group A (VR)	Group B (Ex-vivo)
Number of polyps performed	5	13
Polyp size	9.40mm (SE 1.50)	7.23mm (SE 1.03)
Position of polyp		
Recto-sigmoid	2	8
Descending	1	1
Transverse	1	1
Ascending	1	1
Not recorded	0	2

Table 35: Inter-group comparison for "real-life" polypectomies performed

Number of procedures	Group A /VR	Group B / Ex vivo
Number of colonoscopy lists attended	3.00 (2.00-5.00)	6.00 (0.00-9.50)
Number of OGD lists attended	0.00 (0.00-8.00)	3.00 (0.00-5.50)
Number of Colonoscopies performed	6.00 (5.00-9.00)	15.00 (1.25-29.00)
No. of flexi sigmoidoscopies performed	7.00 (3.00-20.00)	15.50 (1.25-37.00)
Number of OGD's performed	2.00 (1.00-30.00)	12.50 (1.25-37.50)
Number of polypectomies performed	0.00 (0.00-1.00)	3.50 (0.25-4.00)

Table 36: Number of procedures performed from 3rd week training to 4 month follow up

ii. The difference between simulator DOPyS scores for Group A and Group B[§§§]

The first polypectomy performed during each of the 3 training sessions was video recorded and independently DOPyS assessed. It was expected that this would demonstrate an improvement in DOPyS parameters as experience of simulator training increased across the three weeks. **Figure 38** compares the overall pattern of simulator DOPyS scores for each group for each training session. There is no general increase in simulator DOPyS scores across the 3 sessions in either group.

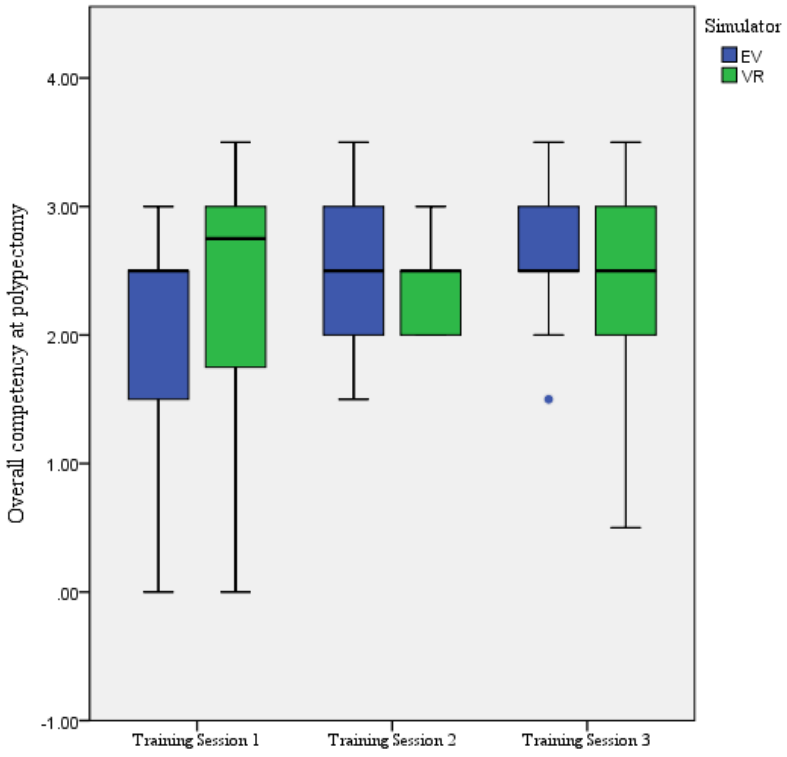


Figure 37: Box plot comparing the overall competency at polypectomy DOPyS parameter scores across training sessions between for Group A / VR and Group B Ex-vivo (EV)

The median simulator DOPyS scores over the 3 sessions of training were comparable between Group A and B (**Table 36**) as were the simulator DOPyS scores per session (**Tables 37, 38 and 39**).

[§§§] Full results can be found in **Appendix 4k-4o**

DOPyS assessment parameter ***All sessions***	Group A / VR	Group B / Ex-vivo
Attempts to achieve optimal position	2.5 (1.5-3.0)	2.5 (2.0-3.0)
Optimises view by aspiration insufflation/ wash	2.5 (2.0-2.9)	2.5 (2.0-3.0)
Directs snare accurately over polyp head	2.5 (2.0-2.9)	2.5 (2.0-3.0)
Places the snare at appropriate position on the stalk	2.5 (2.0-3.0)	2.5 (2.0-3.0)
Ensures appropriate amount of tissue trapped	2.8 (2.0-3.0)	3.0 (2.0-3.0)
Applies appropriate degree of diathermy	2.8 (2.5-3.0)	2.5 (2.5-3.0)
Retrieves or attempts retrieval of polyp	2.5 (2.0-3.0)	2.5 (2.0-2.5)
Overall competency at polypectomy	**2.5 (2.0-3.0)**	**2.5 (2.0-3.0)**

Table 37: Comparison of the median (IQR) simulator DOPyS scores for Group A / VR and Group B / Ex-vivo groups for **all training Sessions**

DOPyS assessment parameter ***Session 1***	Group A / VR	Group B / Ex-vivo
Attempts to achieve optimal position	2.5 (1.8-3.4)	2.5 (2.0-3.5)
Optimises view by aspiration insufflation/ wash	2.5 (1.8-3.3)	2.0 (2.0-3.0)
Directs snare accurately over polyp head	2.3 (1.6-2.9)	2.5 (1.5-2.8)
Places the snare at appropriate position on the stalk	2.5 (1.4-3.0)	3.0 (1.5-3.0)
Ensures appropriate amount of tissue trapped	3.0 (1.4-3.0)	3.0 (1.8-3.0)
Applies appropriate degree of diathermy	2.5 (1.8-3.0)	3.0 (2.0-3.0)
Retrieves or attempts retrieval of polyp	2.8 (1.3-3.4)	2.0 (1.3-2.8)
Overall competency at polypectomy	**2.8 (1.6-3.0)**	**2.5 (1.5-2.8)**

Table 38: Comparison of the median (IQR) simulator DOPyS scores for Group A / VR and Group B / Ex-vivo groups for **Session 1**

DOPyS assessment parameter ***Session 2***	Group A / VR	Group B / Ex-vivo
Attempts to achieve optimal position	2.0 (1.5-2.5)	2.5 (1.5-3.3)
Optimises view by aspiration insufflation/ wash	2.5 (2.0-2.9)	2.5 (2.0-3.0)
Directs snare accurately over polyp head	2.5 (2.1-2.9)	2.5 (2.0-2.8)
Places the snare at appropriate position on the stalk	2.5 (2.0-2.5)	3.0 (2.0-3.0)
Ensures appropriate amount of tissue trapped	2.8 (2.5-3.0)	2.5 (2.0-3.0)
Applies appropriate degree of diathermy	3.0 (2.5-3.0)	2.5 (2.3-2.5)
Retrieves or attempts retrieval of polyp	2.5 (0.9-2.9)	2.0 (2.0-2.8)
Overall competency at polypectomy	**2.5 (2.0-2.5)**	**2.5 (2.0-3.0)**

Table 39: Comparison of the median (IQR) simulator DOPyS scores for Group A / VR and Group B / Ex-vivo groups for **Session 2**

DOPyS assessment parameter ***Session 3***	Group A / VR	Group B / Ex-vivo
Attempts to achieve optimal position	3.0 (1.8-3.0)	2.5 (2.0-3.3)
Optimises view by aspiration insufflation/ wash	2.5 (2.0-2.9)	2.5 (2.0-3.0)
Directs snare accurately over polyp head	2.5 (2.1-2.9)	2.5 (2.5-2.8)
Places the snare at appropriate position on the stalk	2.3 (1.6-3.0)	2.5 (2.0-3.0)
Ensures appropriate amount of tissue trapped	2.3 (2.0-3.0)	3.0 (2.5-3.0)
Applies appropriate degree of diathermy	2.8 (2.0-3.0)	2.5 (2.5-3.0)
Retrieves or attempts retrieval of polyp	3.0 (2.1-3.0)	2.5 (2.0-2.8)
Overall competency at polypectomy	**2.8 (2.0-3.0)**	**2.5 (2.3-3.0)**

Table 40: Comparison of the median (IQR) simulator DOPyS scores for Group A / VR and Group B / Ex-vivo groups for **Session 3**

iii. The difference in simulated polypectomy procedure time for Group A and Group B

The time taken to complete the simulate polypectomy tasks was comparable over the 3 training sessions and in general, increased (**Tables 40, 41 and Figure 39**). Group A completed the simulated polypectomy in less time than Group B (**Figure 40**).

ID	Group A / VR			ID	Group B / Ex-vivo		
	1st Session	2nd Session	3rd Session		1st Session	2nd Session	3rd Session
1	8.07	9.13	11.24	2	8.59	12.21	9.53
5	14.10	13.23	10.17	4	10.04	5.31	5.28
11	8.02	4.28	4.01	7	5.06	12.01	5.37
13	3.36	3.00	3.26	8	10.23	8.59	7.23
14	5.39	4.35	3.38	9	13.22	12.36	8.16
15	4.19	11.16	6.35	10	10.40	6.12	22.00
18	3.54	16.23	14.1	16	9.50	12.10	6.06
19	2.29	6.18	3.51	17	9.08	15.23	11.22
				20	11.07	7.07	20.09
All	6.12 (4.20-9.11)	8.49 (5.46-12.22)	7.00 (4.50-10.08)	All	10.08 (8.36-10.00)	10.13 (7.93-2.13)	10.55 (10.36-5.02)

Table 41: Time taken (mins) to complete VR polypectomy task by Group A (Left) and the Ex-vivo task by Group B (Right) over the 3 training sessions. Total scores (All) are mean value with 95% CI.

Total Cohort		
1st Session	*2nd Session*	*3rd Session*
8.00 (6.43-10.02)	**9.33 (7.56-11.26)**	**9.28 (6.57-12.03)**

Table 42: Time taken (mins) to complete polypectomy task for total cohort over the 3 training sessions. Mean values with 95% CI.

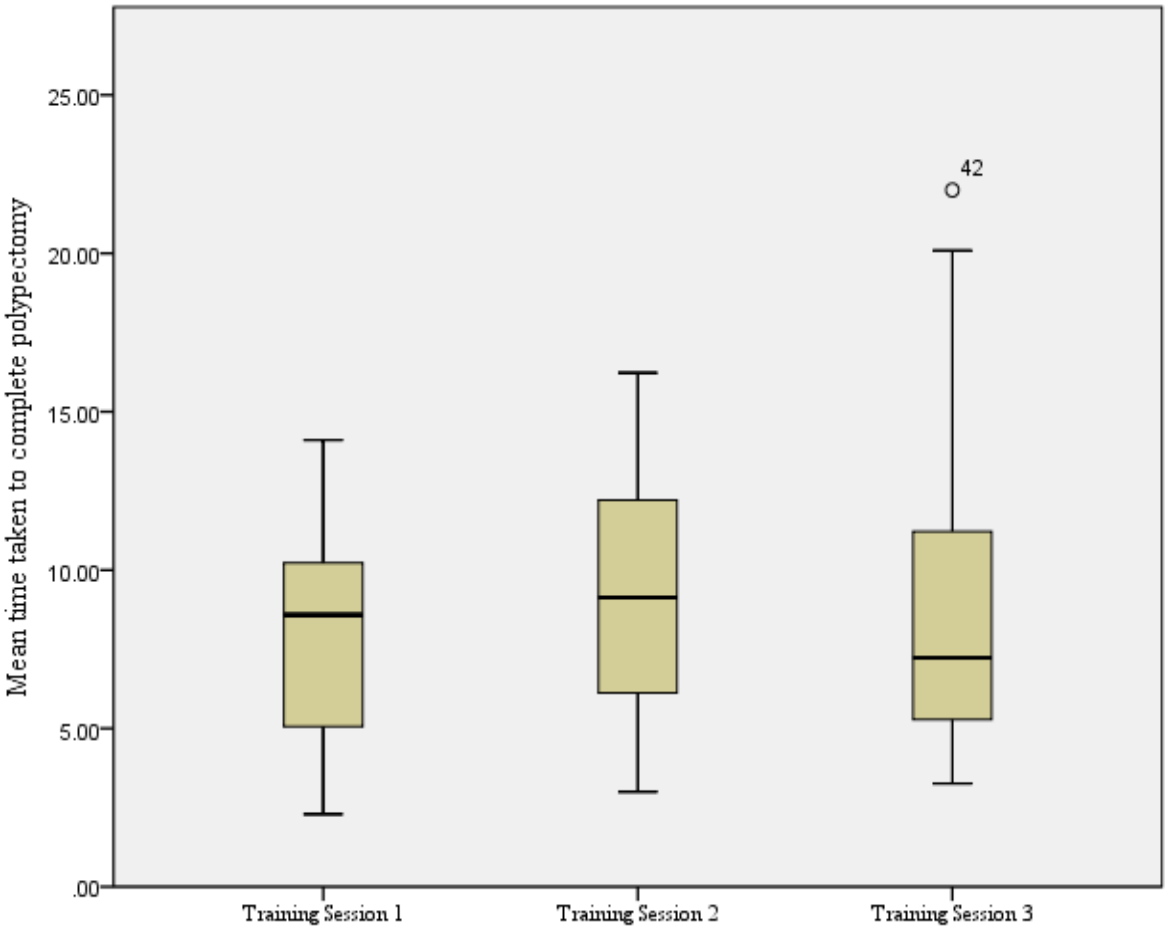


Figure 38: Box plot showing mean time taken (minutes) to complete simulator polypectomy task for each training sessions for total cohort

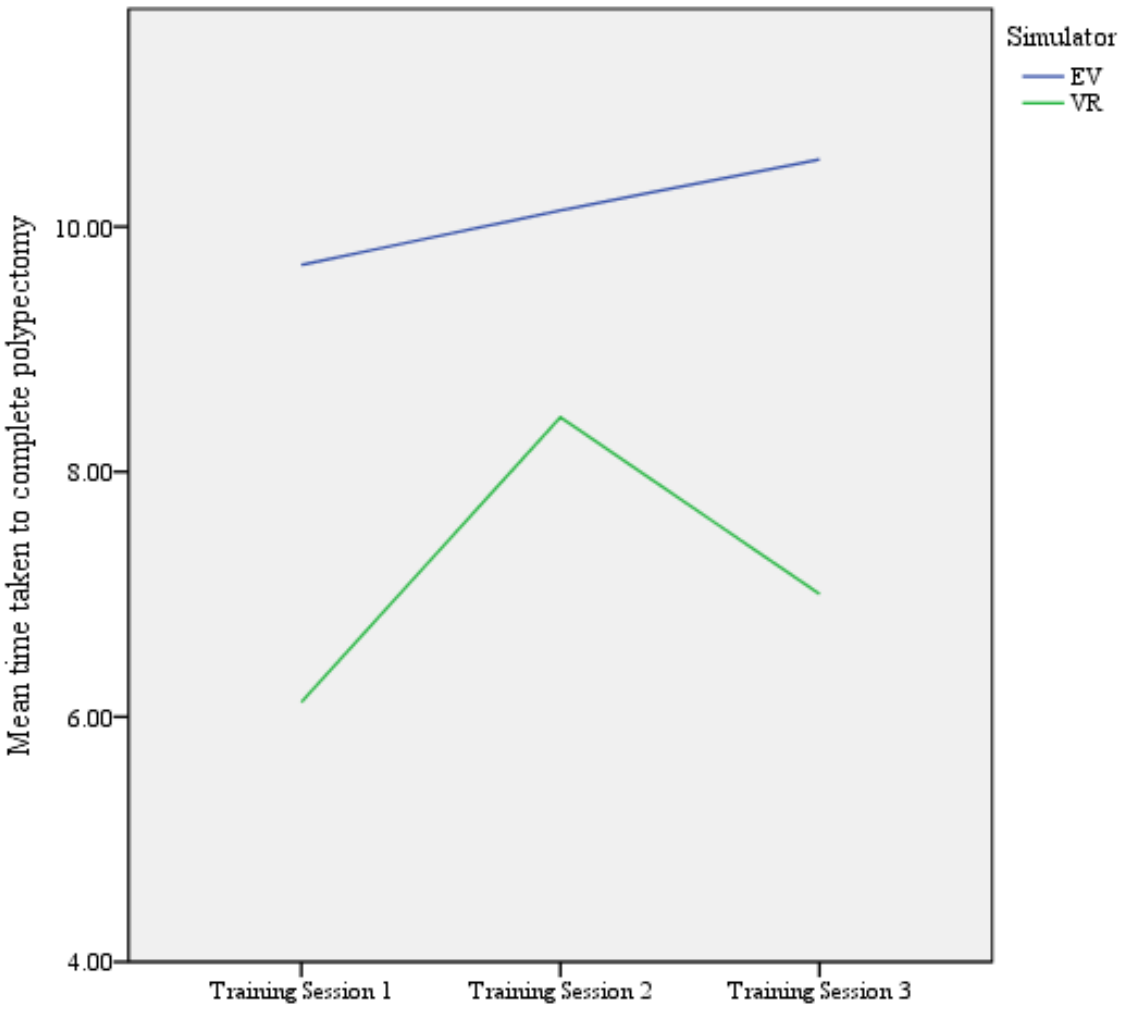


Figure 39: Mean time taken (minutes) to complete simulator polypectomy task for each training sessions for Group A / VR and Group B / EV

iv. The inter-rater correlation between Assessor A and Assessor B

The inter-rater correlation between the 2 independent assessors who rated the simulator videos was overall moderate (**Table 42**) for both the VR (**Figure 41**) and Ex-vivo (**Figure 42**) groups. In general, Assessor B scored both the VR and Ex-vivo participant's marginally higher than Assessor A. There was stronger correlation across all parameters between the 2 assessors scores in the Ex-vivo ($\boldsymbol{\rho}$ = 0.60 (p=<0.01) compared to the VR group ($\boldsymbol{\rho}$ = 0.30 (p=<0.01) (**Table 43**).

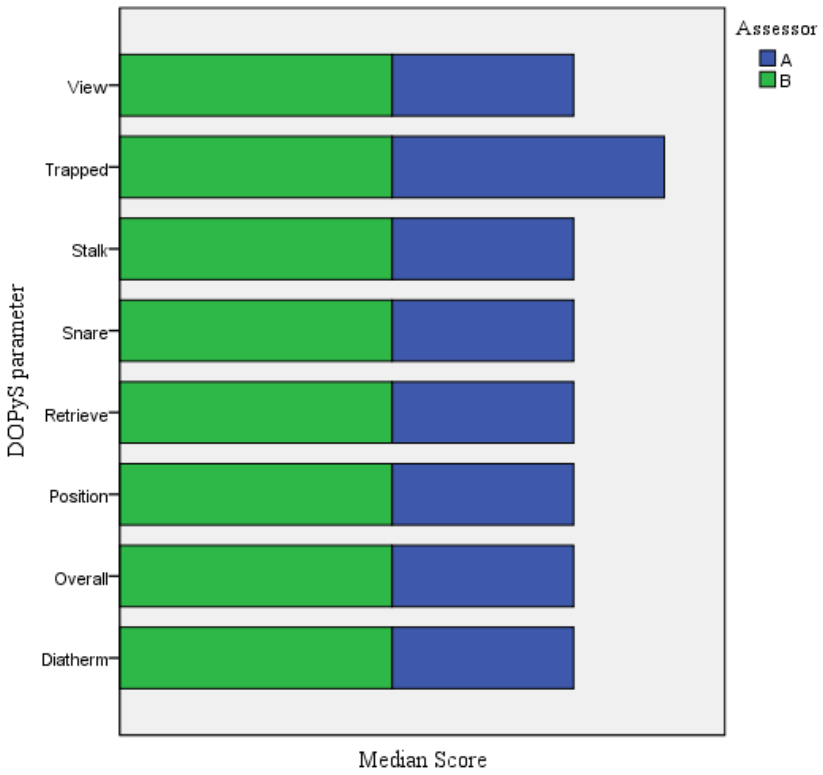


Figure 40: Comparison of Assessor of assessor A and B median DOPyS scores for VR simulator training sessions

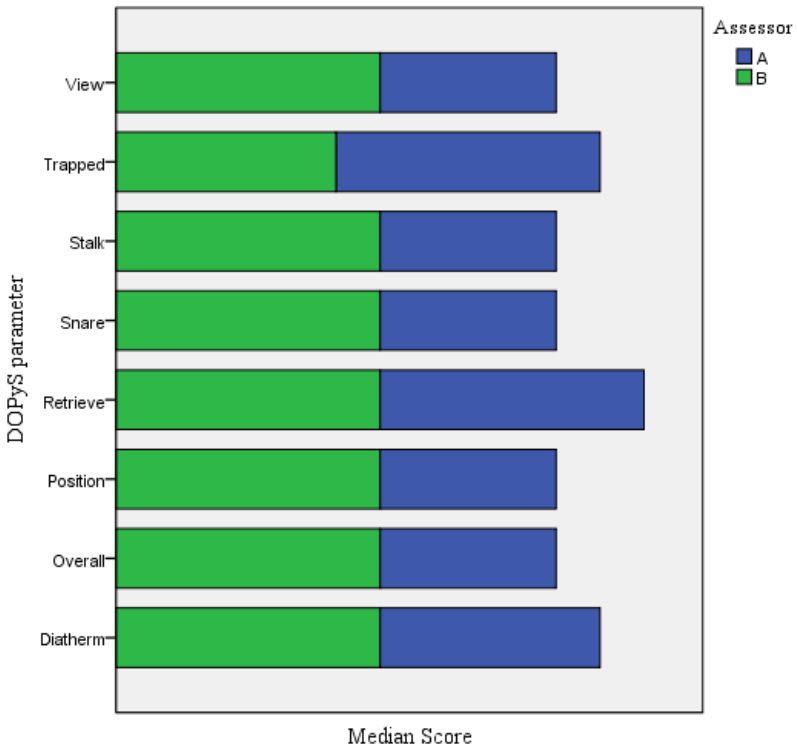


Figure 41: Comparison of Assessor of assessor A and B median DOPyS scores for ex-vivo simulator training

Simulator	Group A	Group B	Correlation
Ex-vivo	2.00 (2.00-3.00)	3.00 (3.00-3.00)	ρ = 0.60 (p=<0.01)
VR	2.00 (1.00-3.00)	3.00 (2.00-3.00)	ρ = 0.30 (p=<0.01)

Table 43: Overall inter-rater correlation for all DOPyS parameters using Spearman's ρ correlation between assessors across the 3 training session where <0.30 was considered to be *weak* correlation, between 0.30-0.50 was considered a *moderate* correlation and >0.50 a *strong* correlation

DOPyS assessment parameter	VR Training			Ex-vivo Training		
	Assessor A DOPyS	Assessor B DOPyS	***Correlation***	Assessor A DOPyS	Assessor B DOPyS	***Correlation***
Attempts to achieve optimal Position	2.0 (1.0-3.0)	3.0 (2.0-3.0)	ρ = 0.65 (p=<0.01)	2.0 (2.0-3.0)	3.0 (2.0-4.0)	ρ = 0.78 (p=<0.01)
Optimises view by aspiration insufflation/ wash	2.0 (1.0-3.0)	3.0 (2.3-3.0)	ρ = 0.50 (p=0.02)	2.0 (2.0-2.0)	3.0 (2.0-4.0)	ρ = 0.41 (p=0.04)
Directs Snare accurately over polyp head	2.0 (1.0-2.8)	3.0 (3.0-3.0)	ρ = 0.61 (p=<0.01)	2.0 (2.0-2.0)	3.0 (3.0-3.0)	ρ = 0.65 (p=<0.01)
Places the snare at appropriate position on the stalk	2.0 (1.0-3.0)	3.0 (2.0-3.0)	ρ = 0.49 (p=0.01)	2.0 (1.0-3.0)	3.0 (2.0-3.0)	ρ = 0.62 (p=<0.01)
Ensures appropriate amount of tissue trapped	3.0 (1.0-3.0)	3.0 (3.0-3.0)	ρ = 0.41 (p=0.05)	3.0 (1.0-3.0)	3.0 (3.0-3.0)	ρ = 0.62 (p=<0.01)
Applies appropriate degree of Diathermy	2.0 (2.0-3.0)	3.0 (3.0-3.0)	ρ = 0.60 (p=<0.01)	2.0 (2.0-2.0)	3.0 (3.0-3.0)	ρ = 0.55 (p=<0.01)
Retrieves or attempts Retrieval of polyp	2.0 (1.3-3.0)	3.0 (2.3-3.0)	ρ = 0.72 (p=<0.01)	2.0 (1.0-2.0)	3.0 (2.0-3.0)	ρ = 0.50 (p=<0.01)
Overall competency at polypectomy	2.0 (1.0-3.0)	3.0 (2.3-3.0)	ρ = 0.45 (p=0.03)	2.0 (1.0-3.0)	2.0 (2.0-3.0)	ρ = 0.67 (p=<0.01)

Table 44: Inter-rater reliability of simulated training sessions using Spearman's ρ correlation between assessors across the 3 training sessions, where <0.30 was considered to be *weak* correlation, between 0.30-0.50 was considered a *moderate* correlation and >0.50 a *strong* correlation.

v. Results from simbionix GI mentor[13]

Additional data was also retrieved from the GI mentor II (Simbionix, USA) for analysis. From previous review work (Chapter 2) there are 4 assessment parameters which have been shown to have construct validity. These are; overall time taken, caecal intubation time, % of the mucosa visualised and the % efficiency of screening performed. Although these parameters are not directly related to polypectomy performance it was interesting to see if the same trend exists across the 3 training sessions for these parameters as the DOPyS scores. **Table 44** and **Figure 43** show the mean scores for all candidates for these parameters for each training session. Like the DOPyS assessment, these demonstrate only marginal improvement in each assessment parameter (**Figures 44 and 45**). Mean scores for individual candidates follow a similar trend (**Table 45**). This further implies that for a larger study we would need to increase either the number of polypectomies performed during each training session or the number of training sessions themselves.

Assessment Parameter	**Mean scores** (95% CI) (Modules 1.5, 1.6, 1.7 & 2.1)		
	1st Session	*2nd Session*	*3rd Session*
Overall time taken (*Minutes*)	11.9 (10.1-13.9)	12.4 (10.3-15.3)	9.8 (8.6-11.2)
Caecal Intubation Time (*Minutes*)	4.2 (3.4-5.1)	4.1 (3.3-4.9)	4.3 (3.1-5.6)
% Mucosa visualised	69.5 (61.8-77.3)	71.6 (63.3-79.7)	73.3 (66.5-79.7)
% efficiency of screening	46.4 (33.8-57.8)	52.4 (40.5-64.5)	58.1 (45.0-69.0)

Table 45: Mean scores generated by the Simbionix GI mentor II (Simbionix, USA) for total cohort across the 3 training sessions

[13] Full results can be found in **Appendix 4p**

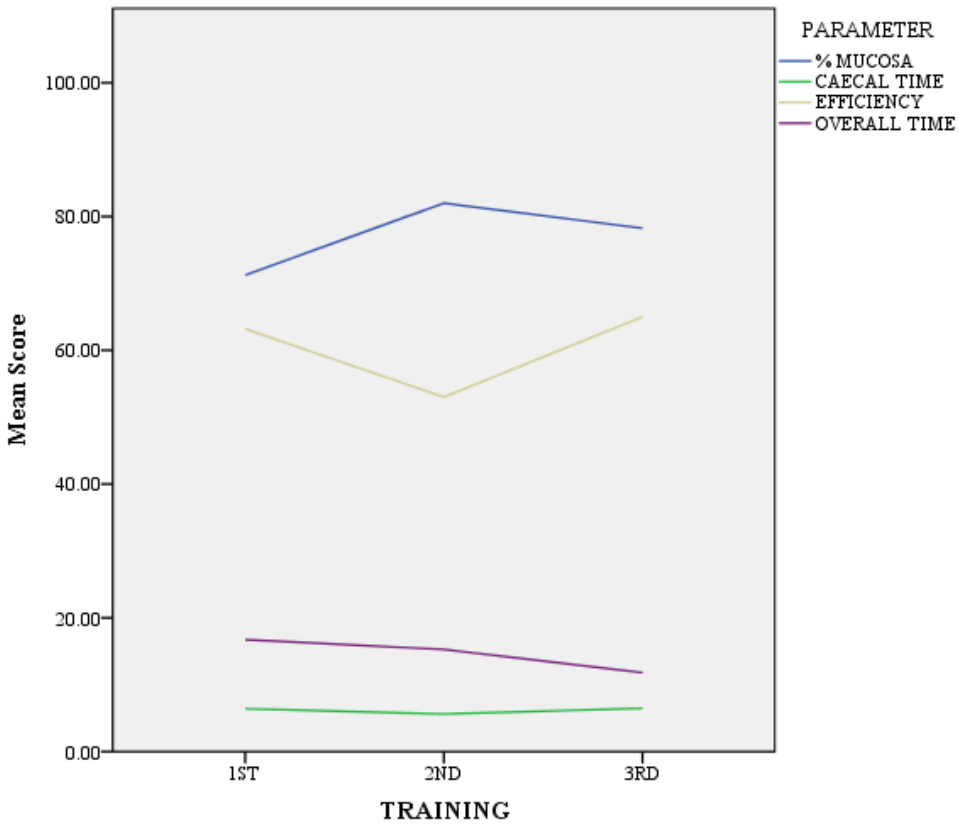


Figure 42: Mean scores generated by the Simbionix GI mentor II (Simbionix, USA) for total cohort across the 3 training sessions for % mucosa visualised, caecal intubation time, % efficiency of screening and overall time to complete the task.

ID	Session	*1st Session*	*2nd Session*	*3rd Session*
1	Overall time taken	16.9 (11.0-22.8)	15.8 (12.1-19.0)	14.4 (11.5-17.2)
	Caecal intubation time	4.8 (2.3-6.1)	4.0 (2.3-5.2)	3.9 (1.6-6.2)
	% Mucosa	78.5 (57.0-92.3)	76.5 (50.5-91.5)	85.5 (80.0-91.0)
	% efficiency	27.5 (13.2-46.5)	35.8 (15.0-54.0)	62.0 (48.8-74.3)
5	Overall time taken	15.3 (6.5-29.3)	12.1 (9.3-15.1)	8.8 (7.6-10.3)
	Caecal intubation time	5.6 (2.2-8.3)	3.9 (2.4-6.2)	3.4 (2.1-5.3)
	% Mucosa	74.3 (57.0-85.8)	85.0 (79.5-90.5)	83.5 (78.0-89.0)
	% efficiency	39.8 (0.0-79.5)	57.3 (20.8-83.5)	64.0 (21.0-88.5)
11	Overall time taken	9.4 (8.6-10.1)	8.5 (6.6-10.3)	8.7 (8.3-9.1)
	Caecal intubation time	3.7 (2.6-4.4)	3.5 (2.0-5.2)	3.1 (2.0-3.9)
	% Mucosa	82.5 (79.0-87.5)	83.8 (78.0-90.3)	83.5 (79.0-88.0)
	% efficiency	83.0 (80.3-85.3)	85.5 (82.0-87.8)	86.5 (81.8-90.5)
13	Overall time taken	11.5 (7.9-15.2)	9.7 (7.4-11.9)	6.9 (3.4-12.0)
	Caecal intubation time	6.2 (6.2-6.2)	6.5 (6.5-6.5)	9.6 (9.6-9.6)
	% Mucosa	79.3 (71.0-88.8)	80.5 (73.0-88.0)	46.0 (21.8-75.8)
	% efficiency	66.3 (55.3-74.5)	76.8 (73.5-80.0)	16.0 (0.0-48.0)
14	Overall time taken	11.0 (6.7-16.0)	9.9 (5.4-13.9)	11.2 (8.6-13.3)
	Caecal intubation time	Not completed	Not completed	4.9 (3.4-6.4)
	% Mucosa	36.5 (26.8-43.2)	36.5 (23.5-47.5)	62.5 (41.5-83.5)
	% efficiency	Not Recorded	Not recorded	36.3 (0.0-72.5)
15	Overall time taken	12.7 (10.3-16.0)	12.3 (8.4-17.6)	11.6 (6.6-11.0)
	Caecal intubation time	3.5 (2.8-4.2)	4.2 (2.3-6.1)	2.8 91.6-4.0)
	% Mucosa	86.5 (83.5-89.5)	84.8 (78.5-91.0)	77.25 (73.0-81.3)
	% efficiency	71.3 (54.8-82.5)	68.8 (45.5-81.8)	76.8 (71.8-81.0)
18	Overall time taken	9.7 (6.2-12.7)	15.0 (11.2-21.4)	11.9 (9.3-16.8)
	Caecal intubation time	Not completed	4.9 (3.2-6.2)	5.8 (3.0-10.1)
	% Mucosa	37.0 (22.3-51.2)	85.5 (81.5-90.3)	85.0 (79.0-92.0)
	% efficiency	Not Recorded	60.5 (35.8-77.5)	72.0 (51.8-85.3)
19	Overall time taken	8.6 (6.8-11.1)	15.8 (2.3-34.7)	7.8 (4.5-11.7)
	Caecal intubation time	4.1 (4.1-4.1)	4.3 (4.3-4.3)	3.3 (3.3-3.3)
	% Mucosa	85.5 (76.8-86.5)	40.0 (8.5-71.5)	63.5 (46.3-74.8)
	% efficiency	83.3 (77.0-89.5)	34.8 (0.0-69.8)	51.5 (20.0-79.0)

Table 46: Mean (95% CI) scores generated by the Simbionix GI mentor II (Simbionix, USA) for individuals across the 3 training sessions

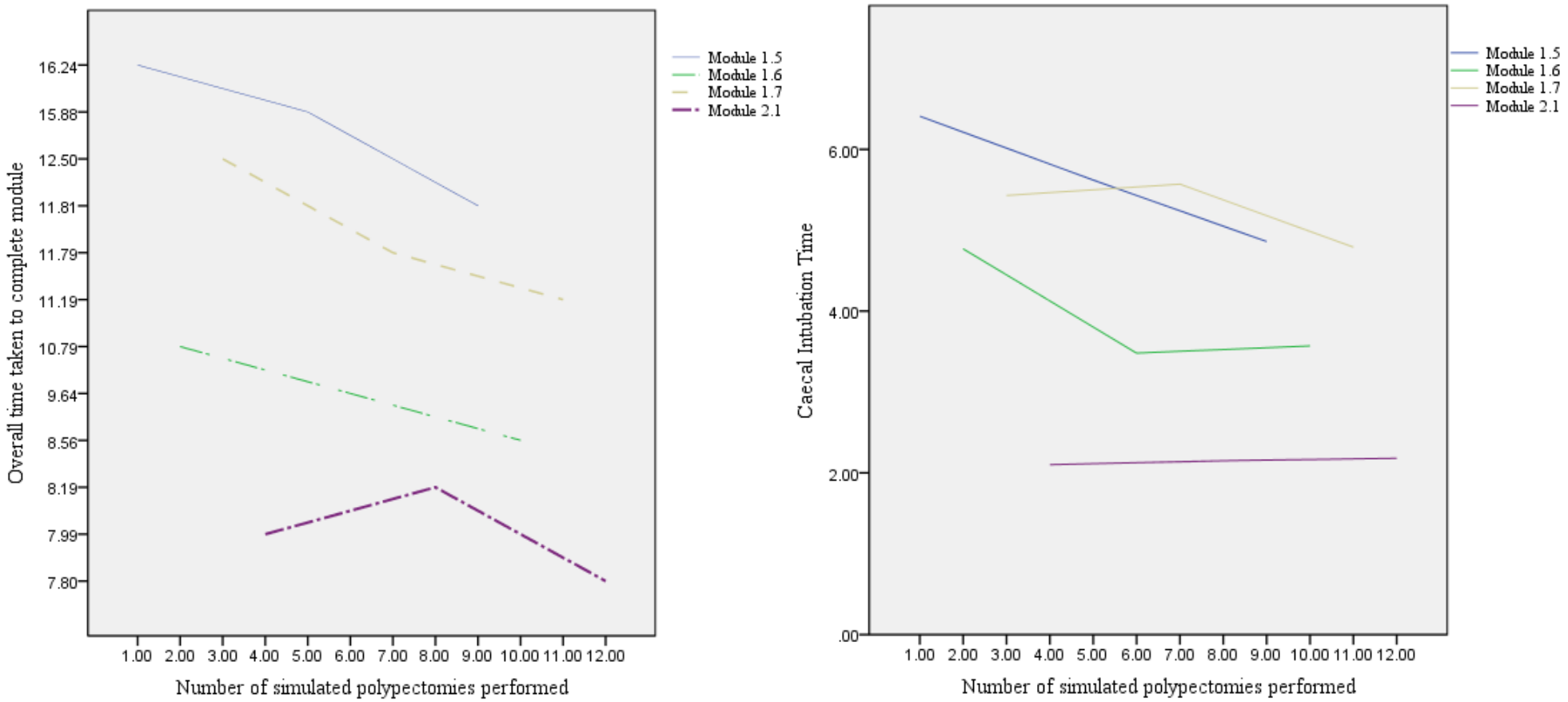


Figure 43: Results from Simbionix GI mentor II VR simulator. Overall time taken to complete each VR module versus number of simulated polypectomies performed (left) and Caecal Intubation time for each VR module versus number of simulated polypectomies performed (right).

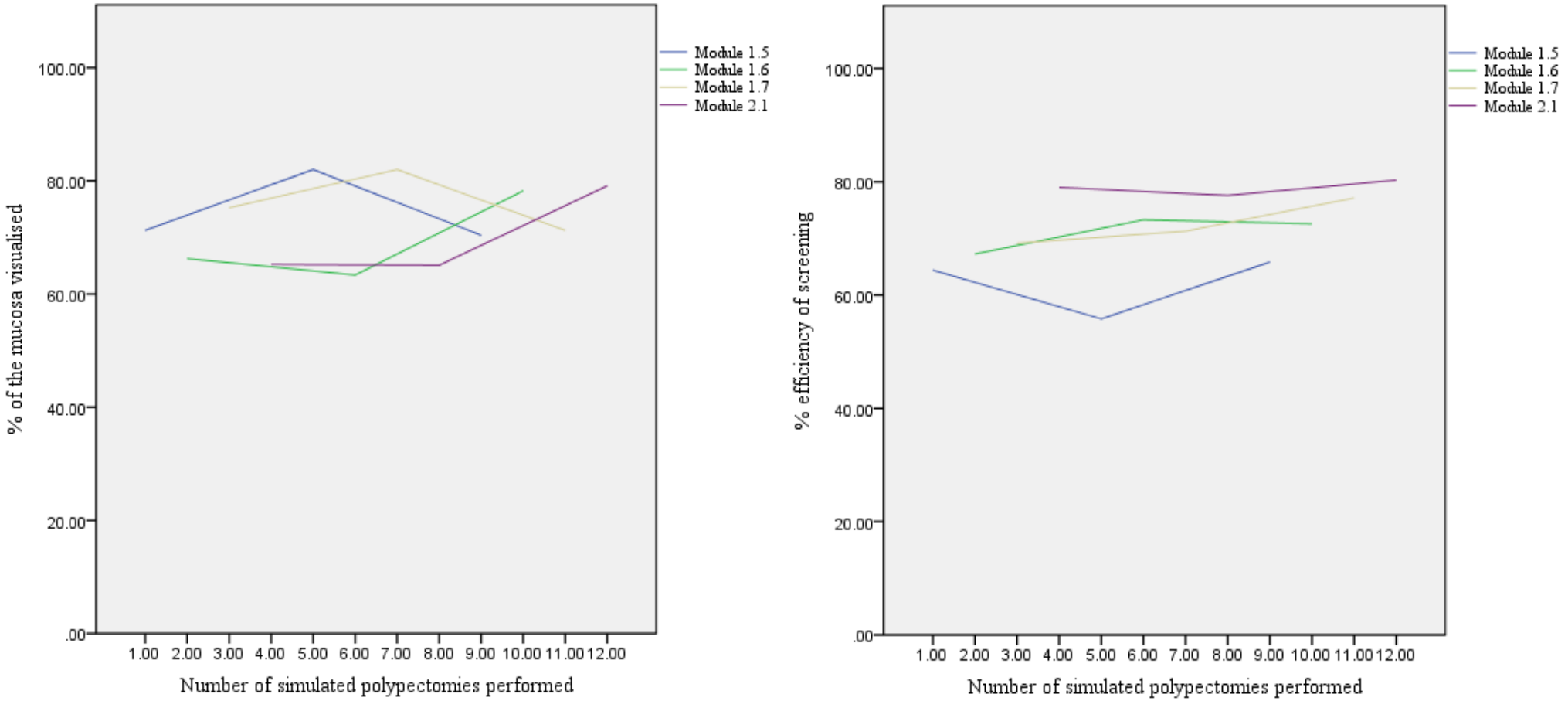


Figure 44: Results from Simbionix GI mentor II VR simulator. Percentage of the mucosa visualised during each VR module versus number of simulated polypectomies performed (left) and percentage efficiency screening during each VR module versus number of simulated polypectomies performed (right)

e) Feasibility outcomes

i. The feasibility of recruitment and the suitability of inclusion and exclusion criteria

It was estimated that 25 trainees across Wales were eligible for inclusion in the trial. The recruitment of 20 participants (10 per group) according to the pre-defined inclusion and exclusion criteria was successful. However, it was not feasible to recruit replacement participants for those who did not respond to the post intervention endpoint "real-life" DOPyS survey or those who were unable to complete any endpoint "real-life" DOPyS assessments within the 4 month follow-up period.

ii. Participant compliance with simulation training regimes

There was good compliance with simulation training regimes. Only 2 participants in Group A and 1 participant in Group B were unable to complete the 3 sessions within the time allocated for simulation training.

iii. The amount of simulation training required to demonstrate skills transfer

It was possible to demonstrate some element of skills transfer following simulation training for those candidates who were able to complete the endpoint "real-life" DOPyS assessment. This was more noticeable in Group B compared to Group A (**Table 33**). It is likely, that in a larger trial, either more simulated cases or increasing training hours would be required to give a more convincing skills transfer affect.

iv. The feasibility of primary and secondary endpoints

The DOPyS score was used in this pilot study as the primary outcome measure of skills transfer. The DOPyS is currently the only validated form of polypectomy skills assessment

(Gupta et al., 2012). Where candidates were able to complete the endpoint process, it was possible to compare the difference in results with the baseline scores. In order to improve the submission of endpoint DOPyS scores in a larger trial, it may be useful change the post simulator training assessment process. The variables which most influenced the submission of endpoint DOPyS scores were the number of training lists attended and the number of endoscopic procedures performed (**Figure 46**).

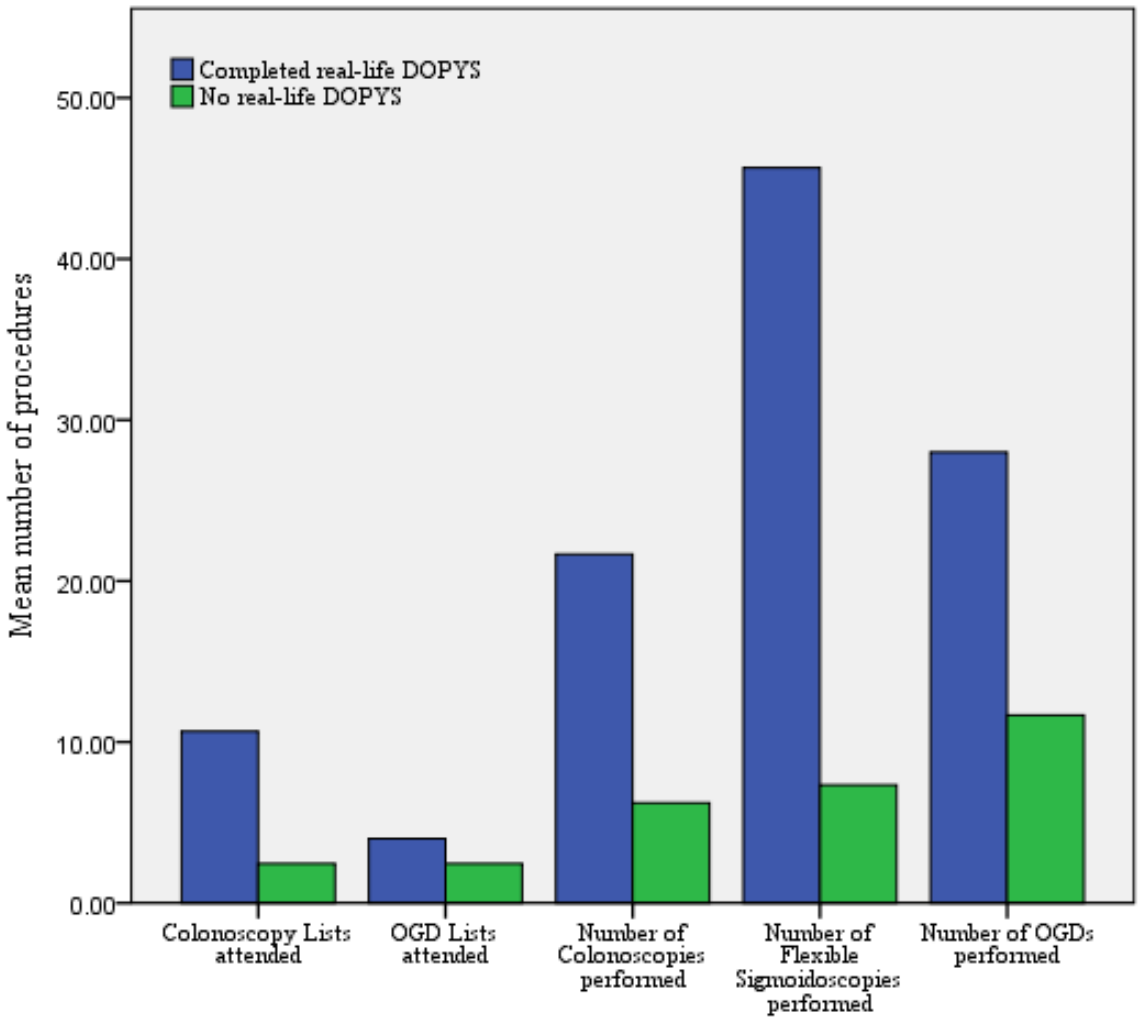


Figure 45: Difference in endoscopic exposure at 4 months post simulator training for those who successfully completed endpoint "real-life" DOPyS scores versus those that did not

Secondary outcomes mainly focussed on comparing the differences in DOPyS scoring between Groups A and B. It is feasible to compare DOPyS scores in this way both in the "real-life" and simulated setting. Standardising "real-life" polypectomy was attempted according to size and morphology of the polyp. All "real-life" polypectomies performed were snared pedunculated polyps, the majority of which were <1cm in size and located in the recto-sigmoid (**Table 32**). For simulated DOPyS scores, inter-rater reliability was strong, further validating the accuracy of this form of scoring.

6.4 Discussion

This pilot trial has demonstrated several important results. Firstly, it was feasible to recruit up to 20 participants from a single deanery despite the fact that the inclusion and exclusion criteria were specific to a small cohort of gastroenterology and surgical trainees. For a larger trial of this nature, it may be necessary use a multi-centre approach in order to recruit further participants to account for drop-out rate and any losses to follow-up. Further engagement at the deanery level to establish mandated simulator training may also prove useful.

a) Improving the measurement of the primary outcome

The main finding and limitation of this pilot was the low rate of endpoint "real-life" DOPyS submission at 4 months follow-up. Firstly, this meant that there were not sufficient results to accurately compare baseline and endpoint "real-life" DOPyS scores. The subgroup analysis of those who did complete "real-life" polypectomies was of a small cohort but a measurable difference was demonstrated between baseline and endpoint scores for both the total cohort and between the groups. The post-simulation training survey identified that exposure to endoscopic training lists was the main factor that determined endpoint DOPyS submission (**Figure 46**). If a larger trial is to be successful, then clearly, access to colonoscopic training

lists for participants is vital. Formalisation of this assessment process may also help to improve endpoint "real-life" DOPyS submission. Following completion of simulator training a formal assessment date could be scheduled with a JAG accredited assessor on a designated endoscopy list. This would standardise the assessment and may improve engagement in the process. However, this would not guarantee that a suitable polyp would arise on the training list for the participant to remove. Recruiting a Principal Investigator at each site to may also help monitor the trial, highlight those participants enrolled and ensure that they have adequate opportunity to perform the necessary number of "real-life" polypectomies.

b) Optimising simulator performance

No significant improvement was demonstrated for simulator DOPyS scores across the 3 training sessions (**Figures 43 and Table 45**). This may be because the shortened DOPyS used was not directly transferrable to the simulator setting. However, this assessment was validated by the strong inter-rater reliability between Assessor A and B (**Table 43**). It was more likely that candidates did not have sufficient training time on either simulator in order to demonstrate a learning curve. The lack of previous research detailing the number of colonoscopic polypectomies needed to reach proficiency made it difficult to establish how many training polypectomies should be performed during this simulator trial to enable skills transfer. It is however, clear from these results that 12 polypectomies over 3 training sessions may not be adequate. Kato *et al* have recently published a paper which focuses on the learning curve of ESD for 150 simulated early gastric cancers (Kato et al., 2013). The authors found that after 30 simulated ESD cases, the two novices performed ESD with a 100% en-bloc resection rate and without perforation (Kato et al., 2013). The total procedure time and perforation rate in the last 30 cases were significantly lower than during the first 30 cases (Kato et al., 2013). Similarly, Sakamoto *et al* have shown that ≥30 cases are sufficient to

perform colorectal ESD without serious complications under the guidance of experienced specialists (Sakamoto et al., 2011). Although there remains no evidence to address the learning curve in simple pedunculated polypectomy, it is likely that 30 simulated cases would be adequate to assess the effect of skills transfer given that this is possible in more complex procedures such as ESD. This increase in cases from 12 to 30 would be incorporated into any future trial.

c) Conclusions

- Evidence of skills transfer being demonstrated following simulation training using the WIMAT colonoscopy suitcase
- That the process of comparing baseline and endpoint DOPyS scores is feasible
- That a degree of protocol revision is needed to ensure that endpoint DOPyS scores are obtainable by improving access to colonoscopy training lists
- The process of video recorded, retrospective analysis of simulator performance using shortened DOPyS is accurate and feasible
- A multicentre approach may be required to ensure that recruitment is feasible

Part 2:

Assessment using the WIMAT colonoscopy suitcase

Chapter 7: Assessment

Can endoscopists accurately self-assess performance during simulated colonoscopic polypectomy?

7.1 Introduction

a) *Ex-vivo simulation and assessment*

The lack of automated assessment parameters for ex-vivo simulation is one of the major limitations of these models. VR simulators have inbuilt parameters which automatically measure performance during a simulation task. This data is presented to the trainee following the simulation to help monitor progression. The trainee can therefore use the simulator independently and review their own performance without the need for an assessor to be present. If ex-vivo simulation is to become as widely utilised as VR, then different forms of assessment need to be explored. One potential method is through self-assessment.

b) Self-assessment

Self-assessment is the ability of an individual to judge their own capabilities and limitations (Pandey et al., 2008). This is a valuable trait for doctors throughout training and independent practice (Pandey et al., 2008, Fitzgerald et al., 2000). Self-assessment however, remains an underdeveloped skill in surgical training and receives little attention from surgical educators (Harrington et al., 1997). It is particularly relevant at present, in view of the General Medical Council's re-validation process (Ball et al., 2011). This requires an ability to recognise strengths and weaknesses to formulate appropriate learning plans. Self-assessment can also impact upon patient care, as either overconfident or unconsciously incompetent surgeons have been shown to jeopardise patient safety (Maslow, 1987).

c) Limitations of self-assessment

Despite the proposed benefit, several studies from a variety of disciplines indicate that the accuracy of self-assessment is poor (Gordon, 1991, Kegel-Flom, 1975, Morton and Macbeth, 1977, Stuart et al., 1980). The majority of research however, focuses on written and clinical knowledge, rather than technical skills. The evidence is contradictory when practical tasks are considered. Moorthy *et al* found that senior surgical trainees are accurate in their self-assessment of technical procedures in a simulated operating theatre whereas, Va Pandey *et al* showed poor correlation between self and expert appraisal for similar skills (Moorthy et al., 2006, Pandey et al., 2008). Other papers report that high self-belief does not predict success and in novices it corresponds negatively with skill (Maschuw et al., 2008).

d) Self-assessment in endoscopy

JAG recommends that trainees use a personal development plan (PDP) in order to highlight learning needs (JAG, 2011). The success of this process is dependent on self-awareness of ability and in turn, self-appraisal. For diagnostic colonoscopy, Koch *et al* reports that self-assessment using the Rotterdam Assessment Form for colonoscopy (RAF-c) can be used to optimise caecal intubation rate. (Koch et al., 2012). The benefit of self-assessment has however, yet to be determined for therapeutic colonoscopy.

e) Aims

The aim of this study was to determine if endoscopists of differing levels of expertise can accurately self-assess their performance in polypectomy using the WIMAT colonoscopy suitcase.

7.2 Methods

a) Study procedure

The self-assessment data presented in this chapter was collected during the construct and concurrent validation trial (**Chapter 5:** *Testing the construct and validity of the WIMAT colonoscopy suitcase,* pages 91-106). During this study 80 participants[††††] completed a standardised polypectomy task. The WIMAT colonoscopy suitcase was used to perform 2 snare polypectomy procedures. The exact positioning of the polyps and construction of the simulator is described in **Chapter 5** of this thesis. The task was designed to allow participants to perform a simple polypectomy (polyp A) and a complex polypectomy (polyp B). The set-up of each simulation was identical and the same endoscopy assistant was present for all cases. The luminal view of all procedures was video recorded, as previously described for retrospective performance analysis.

b) Assessment

Following the polypectomy task, each participant was asked to perform a self-assessment of their performance. This was undertaken using the shortened DOPyS form (**Table 16**). Eight of 34 DOPyS parameters were selected according to their relevance to generic skills and for stalked polypectomy. This generated a "self-assessment" score for the polyp task. Two JAG accredited colonoscopists then reviewed each video and scored the performance using the same DOPyS assessment. Assessors remained blind to the level of experience of the participant at all times. This generated an "expert" score for polyps A and B.

[††††] *Novices* (limited experience of colonoscopy), *Intermediates* (ST3-7), *Advanced* (JAG accredited), Experts (bowel screening colonoscopists).

c) Statistical analyses

The sample size was calculated based on the primary outcome of the construct validation trial. This outcome was the difference in expert DOPyS scores between groups of differing expertise. Previous research has demonstrated that experts (equivalent to our definition of experts) have an 88% (n=15) chance of scoring 3-4 (or pass) on the DOPyS, whereas non-experts (equivalent to our definition of intermediate level) have a 53% (n=8) chance of scoring between 3-4 (or pass) on the DOPyS. Assuming that the novice group would take a similar drop in performance (from 53% to 18%), then 20 participants in each group would give >80% power to detect a difference in performance of 35% between the groups, using a Two-tailed test with a confidence level of 5%.

All statistical analyses were performed on the PASW Statistics Package 18. An expert score for each of the eight DOPyS parameters was calculated by averaging the score of polyp A and B across both assessors. Inter-rater reliability of assessors and the relationship between self-assessment and expert assessment was determined using Spearman's ρ correlation coefficients, where <0.30 was considered to be *weak* correlation, between 0.30-0.50 was considered a *moderate* correlation and >0.50 a *strong* correlation. To complement this analysis a Wilcoxon signed-rank test was used to assess differences in median group performance. Data is expressed as medians with an Inter Quartile Range (IQR) where a *p* value of <0.05 was considered statistically significant.

7.3 Results

a) Group Comparisons

Eighty participants (20 per group) attempted the polypectomy task and all completed the self-assessment process following the simulation (**Table 46**). A statistically significant difference

was demonstrated between the groups for median overall competency assessment scores (novices vs intermediates p=<0.01, novices vs advanced p=<0.01, novices vs experts p=<0.01, intermediates vs advanced p=0.02, intermediates vs experts p=<0.01). No statistically significant difference was seen between advanced vs experts p=0.14). This was also the case for self-assessment median group scores for overall competency (novices vs intermediates p=<0.01, novices vs advanced p=<0.01, novices vs experts p=<0.01, intermediates vs advanced p=<0.01, intermediates vs experts p=<0.01 and advanced vs experts p=<0.01).

b) Self-assessment versus expert scoring

When the expert scores were compared to the self-assessment scores of total cohort the correlation was strong (ρ=0.70 p=<0.01) **(Figure 47)**. However, when the groups were considered separately the correlation was weak. The novice median assessment score for overall competency was 1.00 (0.50-1.25) compared to a self-assessment score of 0.50 (0.00-1.00) (ρ=-0.44, p=0.85). For intermediates, the overall competency assessors score was 2.37 (1.75-2.50) and the self-assessment score 2.00 (1.00-2.00) (ρ=-0.16, p=0.51). In the advanced group, the overall competency assessors score was 2.75 (2.50-3.25) and the self-assessment score 3.00 (3.00-3.00) (ρ=0.16, p=0.50). For the experts, the assessors overall score was 3.00 (2.75-3.44) versus a median self-assessment score of 4.00 (3.00-4.00) (ρ=0.07, p=0.76). A similar correlation was noted for the other 7 DOPyS parameters measured **(Table 47).**

c) Interrater reliability

Interrater reliability between assessors showed a moderate to strong correlation for overall competency in polypectomy; novices (polyp A ρ=0.62 p=<0.01 and polyp B ρ=0.50 p=0.03), intermediates (polyp A ρ=0.30 p=0.30 and polyp B ρ=0.60 p=<0.01), advanced (polyp A

ρ=0.50 p=0.03 and polyp B ρ=0.62 p=<0.01) and experts (polyp A ρ =0.74 p=<0.01 and polyp B ρ =0.30 p=0.21). When an overall group comparison was performed, the correlation was strong for polyp A (ρ=0.80 p=<0.01) **(Figure 48)** and polyp B (ρ=0.80 p=<0.01) **(Figure 49).**

Demographics		**Group**				
		N (n=20)	**I (n=20)**	**A (n=20)**	**E (n=20)**	**Overall (n=80)**
Mean age (yrs) (SE)		30.0 (1.28)	33.4 (0.85)	44.7 (1.30)	46.6 (1.33)	38.6 (0.99)
Gender (% Male)		85%	90%	80%	95%	87.5%
Speciality	Medicine	5%	30%	40%	85%	46%
	Surgery	95%	70%	60%	15%	54%
Previous number of colonoscopies	0-50	20	9	0	0	29
	51-100	0	7	0	0	7
	101-150	0	3	0	0	3
	151-200	0	1	0	0	1
	200-500	0	0	3	0	3
	500-1000	0	0	12	16	28
	>1000	0	0	5	4	9
Previous number of pedunculated polypectomies	0-50	20	20	6	6	52
	51-100	0	0	5	8	13
	101-150	0	0	4	1	5
	151-200	0	0	1	1	2
	200-500	0	0	0	0	0
	>500	0	0	4	4	8

Table 47: Demographic data from participants

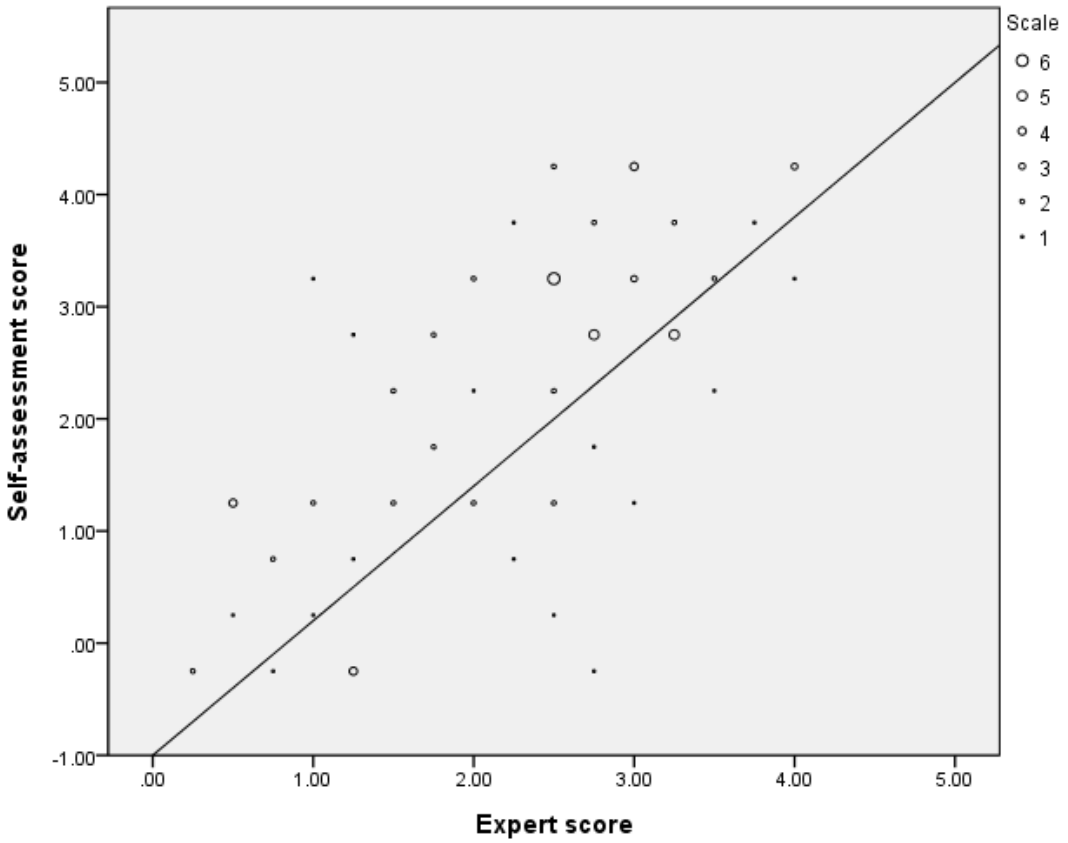


Figure 46: Comparison of expert versus self-assessment scores for overall competency at polypectomy (scale shows binned data where size of the plot indicates the number of data points with identical values)

Group	Score	DOPyS parameter assessed							
		Attempts to achieve optimal position	Optimises view by aspiration/ insufflation	Snare accurately over polyp head	Snare at appropriate position on the stalk	Appropriate amount of tissue trapped	Applies appropriate degree of diathermy	Retrieves or attempts retrieval of polyp	Overall competency at polypectomy
Novices (n=20)	Expert score Median (IQR)	1.00 (0.75-1.25)	1.00 (0.50-1.25)	0.63 (0.63-1.25)	1.00 (0.50-1.25)	1.00 (0.56-1.19)	1.00 (0.50-1.25)	1.00 (0.31-1.25)	1.00 (0.50-1.25)
	Self-assessment Median (IQR)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)
	Spearman's (***ρ***) p value	-0.71 p=0.77	-0.10 p=0.68	0.14 p=0.56	-0.44 p=0.85	-0.12 p=0.61	0.53 p=0.83	-0.07 P=0.77	-0.44 p=0.85
Intermediate (n=20)	Expert score Median (IQR)	2.00 (1.75-2.69)	2.13 (1.75-2.50)	2.00 (1.75-2.50)	2.13 (1.50-2.50)	2.25 (1.75-2.75)	2.13 (1.81-2.50)	2.25 (2.00-2.75)	2.38 (1.75-2.50)
	Self-assessment Median (IQR)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)
	Spearman's (***ρ***) p value	0.06 p=0.82	0.01 p=0.97	-0.20 p=0.39	-0.14 p=0.56	-0.11 p=0.65	-0.23 p=0.33	-0.15 P=0.54	-0.16 p=0.51
Advanced (n=20)	Expert score Median (IQR)	2.87 (2.31-3.25)	2.75 (2.25-3.19)	2.75 (2.31-3.19)	2.75 (2.31-3.25)	3.00 (2.25-3.25)	2.75 (2.31-3.25)	2.50 (2.06-3.19)	2.75 (2.50-3.25)
	Self-assessment Median (IQR)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)	3.0 (3.0-3.0)
	Spearman's (***ρ***) p value	-0.11 p=0.64	0.09 p=0.72	0.18 p=0.46	0.19 p=0.42	0.14 p=0.56	-0.06 p=0.80	-0.10 p=0.68	0.16 p=0.50
Experts (n=20)	Expert score Median (IQR)	3.13 (2.56-3.69)	3.00 (2.50-3.44)	2.75 (2.50-3.50)	3.00 (2.56-3.44)	3.00 (2.56-3.44)	3.00 (2.56-3.44)	3.00 (2.56-3.44)	3.00 (2.75-3.44)
	Self-assessment Median (IQR)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)	4.00 (3.00-4.00)
	Spearman's (***ρ***) p value	0.12 p=0.62	0.16 p=0.50	0.36 p=0.12	0.39 p=0.09	0.22 p=0.34	0.27 p=0.26	0.03 P=0.91	0.07 P=0.76

Table 48: Expert versus self-assessment DOPyS scores for all groups and for each DOPyS parameter used. Spearman's ***ρ*** correlation where <0.30 was considered to be *weak* correlation, between 0.30-0.50 was considered a *moderate* correlation and >0.50 a *strong* correlation.

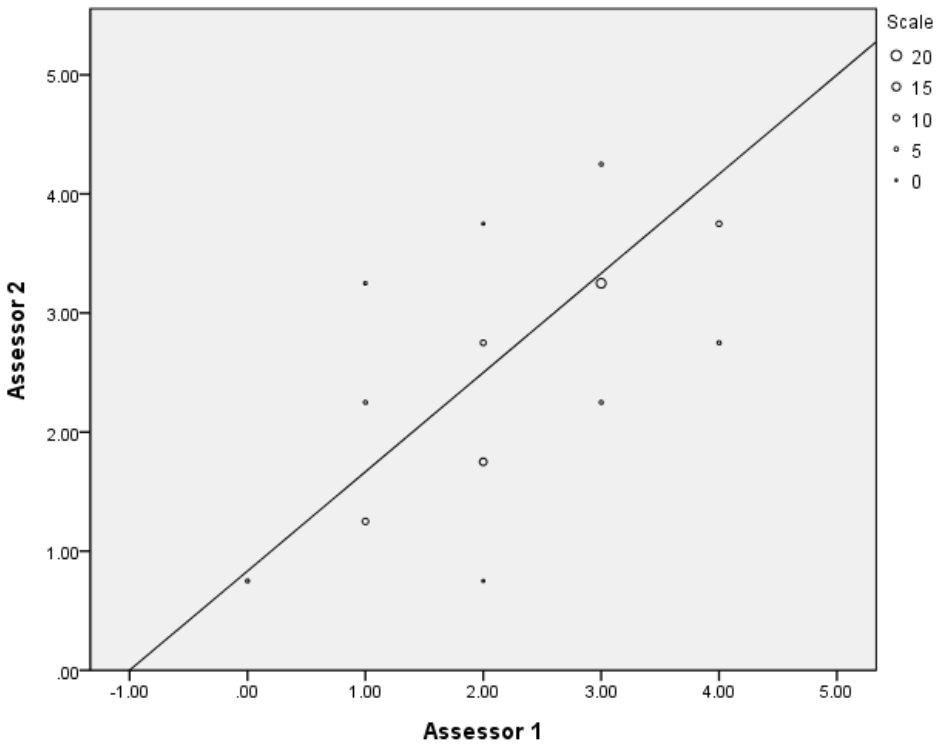


Figure 47: Comparison of assessor 1 and assessor 2 for polyp A (scale shows binned data where size of the plot indicates the number of data points with identical values)

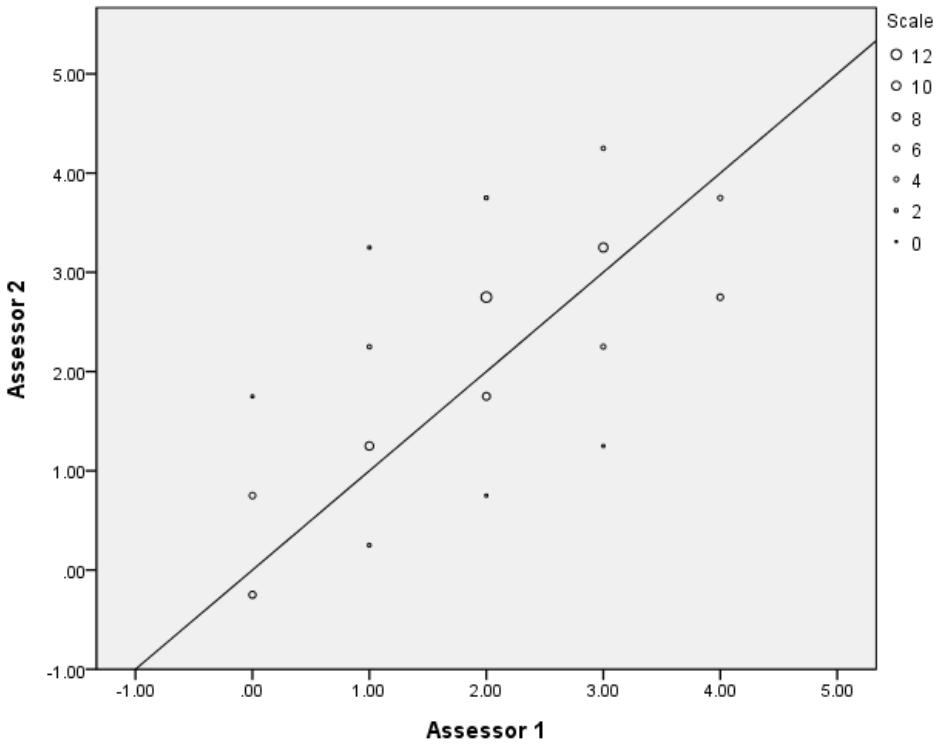


Figure 48: Comparison of assessor 1 and assessor 2 for polyp B (scale shows binned data where size of the plot indicates the number of data points with identical values)

7.4 Discussion

a) Summary of results

This is the first study to evaluate the reliability of self-assessment during simulated colonoscopic polypectomy. There is lack of consistency in the literature to conclude whether accurate self-assessment of technical skill is possible. A meta-analysis of 44 self-assessment studies in higher education reported a moderate correlation between self- and expert-assessments of p=0.39 (Falchikov NB, 1989). A similar review by Gordon *et al* of 18 papers demonstrated comparable findings (Gordon, 1991). However, when medical trainee self-assessments are compared with expert scores, the correlation is usually weak (Pandey et al., 2008, Evans et al., 2007, Davis et al., 2006, Brewster et al., 2008).

Our results for trainee (novices and intermediates) self-assessment in simulated colonoscopic polypectomy are consistent with these findings. The relationship between self-assessment and expert assessment of advanced and expert colonoscopists again, demonstrates weak correlation. This is contrary to some reports in the literature that the ability to accurately self-assess improves with experience because the participant can recognise an expert performance and use this as a benchmark to assess their own skills (Moorthy et al., 2006, Ward et al., 2003).

b) Group comparisons of expert versus self-assessment

Novices and intermediates tended to under estimate their ability (median assessors score for novices 1.00 (0.50-1.25) and 2.38 (1.75-2.50) for intermediates compared with a self-

assessment median score for novices of 0.50 (0.00-1.00) and 2.00 (1.00-2.00) for intermediates. The advanced and experts groups on the other hand overestimated their ability (median assessor score for advanced 2.75 (2.50-3.25) and 3.00 (2.75-3.44) compared with a self-assessment median scores of 3.0 (3.00-3.00) for advanced and 4.00 (3.00-4.00) for experts). These findings can be interpreted in several ways. Firstly, trainees lack experience of performing "real-life" polypectomy and are therefore, less able to accurately self-assess their performance. Advanced and expert colonoscopists may be familiar with assessing novice and intermediate performances but may be less able to repeat this process for themselves. They may also feel pressurised to represent themselves in the best possible light. This is known as *impression management* and has been previously applied to trainees but may in fact, also be true of trainers themselves (Evans et al., 2002). The reinforces the need for continuing professional assessment following endoscopic accreditation.

Alternatively, the over or underestimation can be directed at the assessment process itself. There are currently no valid, quantitative measures of polypectomy assessment. The DOPyS score is a subjective measure of skill that may be open to interpretation. It has been validated for assessment of a range of "real-life" polypectomy procedures but not for simulated polypectomy. This may affect the reliability of its use in this trial. However, 2 blinded, independent JAG accredited assessors demonstrated strong correlation for both polyp tasks which implies that the assessor scores are an in fact, accurate.

Finally, the process of self- and expert-assessment differed. Assessors watched video recordings retrospectively and therefore, had time to scrutinise technique. Self-assessment scores were generated directly following the procedure. Martin *et al* has demonstrated that the correlation between expert and self-assessment improves from 0.38 to 0.52 for communication skills if the participant reviews the video performance rather than from memory (Martin et al., 1998). This effect is diminished in more senior residents and experts

(Martin et al., 1998). It would be interesting to see if this effect was more pronounced in technical skills, self-assessment.

c) Limitations of study

A limitation of this trial may be that a simulator was used in order to replicate a "real-life" scenario. This was chosen in order to standardise the polypectomy task which would be difficult to achieve in the clinical setting. This simulator has undergone validation testing but could never fully replicate colonoscopic polypectomy and therefore, it would be interesting to test self-assessment during a "real-life" scenario in the future.

d) Conclusions

- There is weak correlation between self-assessment and independent expert assessment for simulated colonoscopic polypectomy
- Novices tend to underestimate whereas experts over estimate performance
- Independent expert assessment appears to be a strong way of reliably assessing performance in therapeutic colonoscopy
- Further work is needed in order to develop accurate objective forms of assessment

Chapter 8: Assessment

Rotation To Angulation, (RoTA score): a novel objective performance indicator in therapeutic colonoscopy

8.1 Introduction

a) Objective assessment in lower GI endoscopy

There are no validated objective assessments to determine when an endoscopist has reached technical competence (Lee et al., 2008). However, research is currently being published which puts forward several novel assessment parameters for future development (Obstein et al., 2011, Almansa et al., 2011).

b) Kinematics and motion analysis in colonoscopy

The analysis of motion or kinematics has been used to define surgical expertise in laparoscopic surgery for the last 2 decades. This has led to the development of objective performance measures over a range of laparoscopic tasks (Smith et al., 2002). Colonoscopy has long been neglected in the expanding field of kinematic analysis because the methods and metrics derived for laparoscopy until recently, could not be directly applied to flexible instrument manipulation (Obstein et al., 2011). The first use of motion tracking in endoscopy focussed on Endoscopic Ultrasound (EUS) performance (Vosburgh et al., 2007). The authors used an image registered gastroscopic ultrasound system (IRGUS) to determine EUS performance by kinematic variables (Vosburgh et al., 2007). Subsequent work by Obstein *et al* highlights the potential use of kinematics in order to evaluate technical skill during simulated colonoscopy (Obstein et al., 2011). Electromagnetic sensors were attached to a colonoscope (**Figure 50**) measuring path length, tip-angulation, absolute roll and scope curvature to quantify performance. Differences in these parameters were found according to the expertise of each participant to provide a mean kinetic score for overall performance. Clark *et al* used similar techniques to develop a quantitative scale of endoscopic torque control during Natural Orifice Transluminal Endoscopic Surgery (NOTES) using motion

tracking of wrist movements (Clark et al., 2013). Similar studies which focus on motion analysis have yet to be performed for colonic polypectomy.

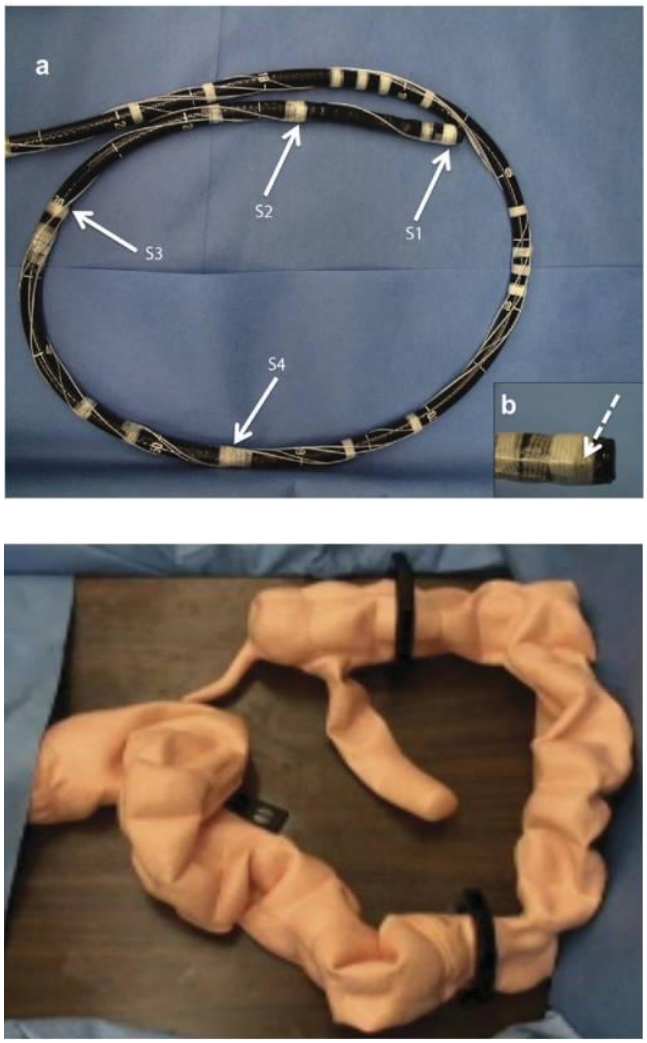


Figure 49: Set-up of kinematic analysis. Top = electromagnetic motion tracking sensors attached to colonoscope, Bottom = colonoscopy bench model simulator (Adam Rouilly, Koken, UK)

Motion analysis technology has also recently been introduced into a novel system for measuring quality in colonoscopy procedures (Endodex, 2013). The EndoDex (EndoDex, UK) allows all routine procedures to be objectively assessed and real time feedback is given

to the clinician using a disposable sensor and image processing algorithms (Endodex, 2013). This technology has good potential but requires further validation before widespread uptake.

c) Eye Gaze Tracking

Eye Gaze Tracking (EGT) is another technology which has the potential to be used in endoscopic performance assessment. This system follows the motion produced by or in a visual gaze pattern of saccadic eye movements and fixations (Almansa et al., 2011). EGT has previously been used in order to explore potential training tools in laparoscopic surgery and has been shown to improve technical skill and multi-task performance (Chetwood et al., 2012, Wilson et al., 2011). It has also been used by radiologists to evaluate and enhance diagnostic yield (Kundel et al., 2007). More recently this technology has been applied to colonoscopy in an attempt to improve quality. Almansa *et al* studied 11 endoscopists watching 3 videos of colonoscopy while their Visual Gaze Pattern (VGP) was recorded (Almansa et al., 2011). The authors recorded a positive correlation between centrally focussed VGP and adenoma detection rate (Almansa et al., 2011).

d) Video tagging software

Another area which has not been previously explored as a measure of performance in endoscopy is video tagging. This is widely utilised in sporting analysis (**Figure 51**) and is gaining popularity in medical team training (**Figure 52**). Video tagging enables capture, categorisation, monitoring and assessment of video recorded procedures. One of the most widely used software tools for this purpose is StudioCode (Sportste™, Australia). This system analyses total performance and provides effective feedback for improving outcomes (Studiocode, 2013). This has not been used in the analysis of surgery or gastroenterology

before. It was therefore, decided to pilot this approach alongside the WIMAT colonoscopy suitcase.

Figure 50: The use of video performance tagging for assessment of team performance in sports (Studiocode Version 4.5.1, Sportstec™) (Studiocode, 2013)

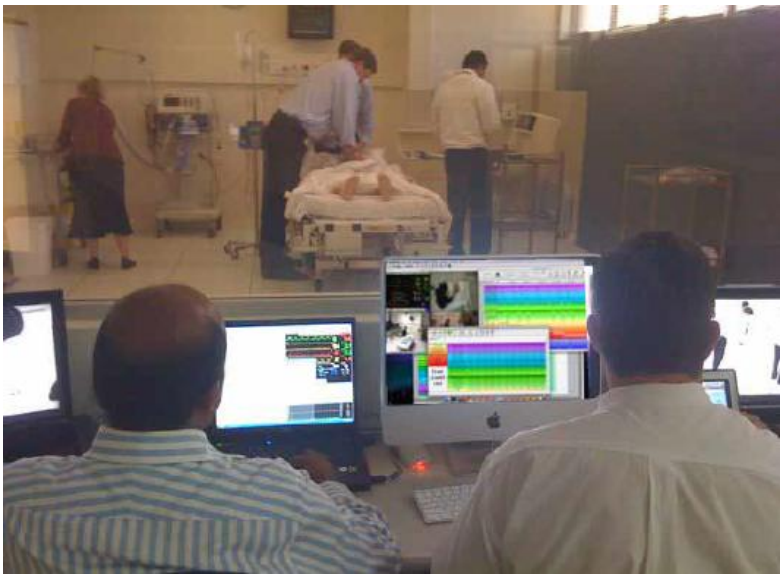


Figure 51: The use of video performance tagging for team training in a simulated environment (Studiocode Version 4.5.1, Sportstec™) (Studiocode, 2013)

e) Video tagging using the WIMAT colonoscopy suitcase

Previous endoscopic assessment measures have focused on the evaluation of the luminal performance. We hypothesized that a potentially important, overlooked area of interest is the hand movements used to perform the task. In colonoscopy, torque steering is achieved by combining rotation of the endoscopic shaft with up or down angulation of the scope tip where the left/right wheel is used as little as possible (GastroTraining, 2013). This is a widely accepted technique in colonoscopy and is taught to novices during JAG training (JAG, 2012). There is however, no evidence which links this method of endoscope control to performance in therapeutic colonoscopy. It is unclear how the ratio between shaft rotation and tip control differs between novices and experts. A proven quantifiable ratio may be useful for objective assessment of performance.

f) Aims and hypotheses

The aim of this study is to use StudioCode (Sportstec^{TM}, Australia) with the WIMAT colonoscopy suitcase to evaluate how hand movements performed during simulated polypectomy correlate with DOPyS scoring. It was felt that this may identify potential objective performance parameters. The following hypotheses were generated:

- There is a difference in the hand movements used by novices and experts when performing simulated polypectomy

- Experts use rotational steering movements (as opposed to tip movements) more than novices when performing simulated polypectomy

8.2 Materials and Methods

a) Design, participants and setting

The data presented in this chapter was collected during the construct and concurrent validation trial (Chapter 5: *Testing the construct and concurrent validity of the WIMAT colonoscopy suitcase*). All participants attended the Welsh Institute for Minimal Access Therapy (WIMAT) between August and December 2012. A preliminary questionnaire was used to identify candidate demographics including previous endoscopic and simulator experience. All procedures took place within the simulated endoscopy suite at the WIMAT**.**

b) Simulation Task

Participants performed a single, standardised stalked polypectomy on the WIMAT colonoscopy suitcase. The previously validated Polyp A position was used as a standardised task to complete. This polyp was 1cm in size (from stalk base to polyp head) and the length of the stalk exposed was 5mm. The polyp was inserted into a straight, standard length of porcine bowel, 25cm from the anal verge in the 6 O'clock position, in front of a luminal fold. The endoscopic equipment used and the task were the same as previously described[‡‡‡‡].

c) Assessment

The endoscopic view from each polypectomy was video recorded and 2 independent, JAG accredited assessors scored each polypectomy using the shortened DOPyS assessment (**Table 16**). Both assessors remained blind to the level of the candidate performing the polypectomy at all times.

[‡‡‡‡] Chapter 5: *Testing the construct and concurrent validity of the WIMAT colonoscopy suitcase*

d) Hand Analysis

All participants had their hand movements video recorded during the simulated polypectomy using a Tablet computer (iPAD 2, Apple, USA) (**Figure 53**). These recordings were then inputted into a performance analysis software package (Studiocode Version 4.5.1, Sportstec^{TM}) (**Figure 54**). All videos were edited to include the exact point when the endoscope was inserted into the simulator to the first time that the excised polyp was visualised outside the simulator. Two assessors blinded to the participant's identity then coded each performance according the following novel outcome measures:

- The number of times clockwise rotation was applied
- The number of times anticlockwise rotation was applied
- The number of times up or down angulation was used (large control wheel)
- The number of times right or left angulation was used (small control wheel)
- The ratio of rotation to angulation (*Rotation To Angulation, RoTA score*)

e) Statistical Analysis

A median of the 2 assessor's DOPyS scores was calculated for each participant to create the following groups; DOPyS score of <3 (*Group 1*) and DOPyS score of ≥3 (*Group 2*). The ratio of the total number of rotational movements to the total number of angulation movements was also recorded. This provided the *Rotation To Angulation (RoTA)* score. A median of the 2 assessors scores for each hand analysis variable was also calculated. Hand analysis scores were compared between *Group 1* and *Group 2* to identify statistically significant variables. Interrater reliability between assessors was calculated using the Spearman's (ρ) rank for DOPyS assessment and Pearson's rank (r) for hand analysis scores. A score of <0.30 was considered to be *weak* correlation, between 0.30-0.50 was considered a

moderate correlation and >0.50 a *strong* correlation. A Mann-Whitney U test was used for group comparisons. All statistical analysis was performed using SPSS version 18.

Figure 52: Recording the hand movements performed during simulated polypectomy using a Tablet computer (iPAD 2, Apple, USA)

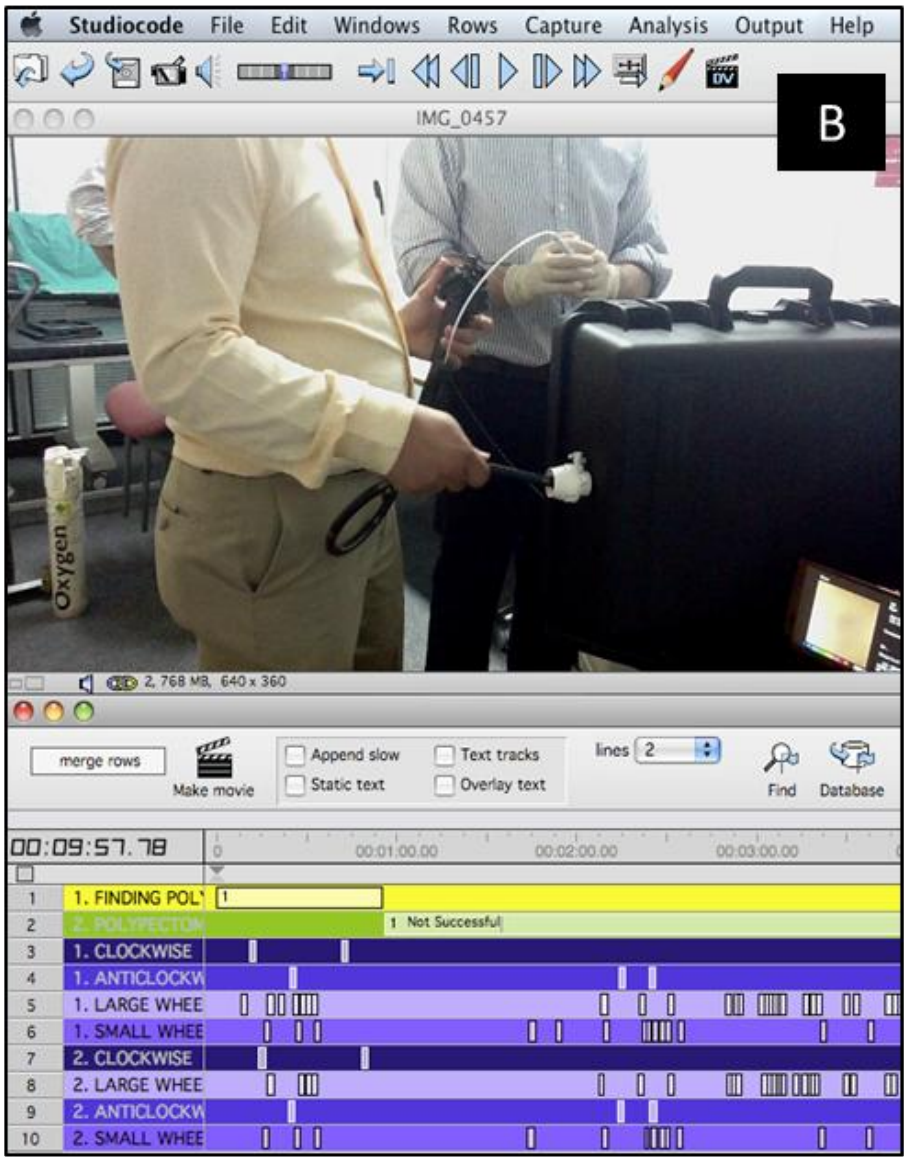


Figure 53: Example of the hand analysis set-up using Studiocode Version 4.5.1, Sportstec, Australia. Hand parameters can be seen below the video image. Each white vertical bar represents a period of time when that movement was used

8.3 Results

Twenty-three participants completed the simulated polypectomy task. Twelve (52%) scored ≤3 on the DOPyS, forming *Group 1* and 11 (48%) >3 on the DOPyS, forming *Group 2*. Statistically significant differences between each group were demonstrated for several demographics including previous experience in endoscopy (**Table 48**).

Hand movements (median IQR)	DOPYS Score		p value
	Group 1 DOPyS <3 (n=12)	*Group 2* DOPyS ≥3 (n=11)	
Age (years)	31.50 (28.25-39.50	43.00 (38.00-48.00)	0.01
Previous number of "real-life" Colonoscopies	10.00 (0.00-50.75)	500.00 (300.00-1100.00)	<0.01
Previous number of "real-life" Flexible Sigmoidoscopies	10.00 (0.00-125.00)	500.00 (400.00-550.00)	<0.01
Previous number of "real-life" Sessile polypectomies	0.00 (0.00-31.25)	60.00 (30.00-100.00)	<0.01
Previous number of "real-life" Pedunculated polypectomies	0.50 (0.00-31.25)	80.00 (50.00-100.00)	<0.01
Median DOPyS for simulated polypectomy	1.75 (1.00-2.38)	3.00 (3.00-3.50)	<0.01

Table 49: Demographics for participants according to DOPyS score; *Group 1* (DOPyS <3) and *Group 2* (DOPyS ≥3). Differences between groups calculated with a Mann-Whitney U test for non-parametric data

a) Hand movement analysis

There were no statistically significant differences found between *Group 1* and *Group 2* for the number of times clockwise rotation was applied, number of times anticlockwise rotation

was applied, number of times up or down angulation was used and the number of right or left angulation was used (**Table 49**). There was a statistically significant difference found between each group for the total time taken to complete the task (p=0.01).

Hand movements (median IQR)	DOPYS Score		p value
	Group 1 DOPyS <3 (n=12)	*Group 2* DOPyS ≥3 (n=11)	
Up or down angulation	57.25 (49.63-95.25)	67.50 (34.00-109.50)	1.00
Right or left angulation	21.75 (6.00-30.50)	24.00 (11.00-45.50)	0.20
Total angulation	84.25 (56.50-113.87)	78.00 (45.50-176.00)	1.00
clockwise rotation	22.50 (11.88-50.25)	19.00 (5.00-32.00)	0.38
anticlockwise rotation	27.25 (12.13-63.00)	9.00 (7.00-30.50)	0.10
Total rotation	50.00 (24.75-106.38)	28.00 (12.00-61.50)	0.21
Time taken (seconds)	456.00 (254.00-705.00)	249.00 (1.98.00-274.00)	0.01
RoTA score	**0.85 (0.40-1.28)**	**0.30 (0.20-0.50)**	**0.04**

Table 50: Comparison of hand movement parameters grouped according to DOPyS score; *Group 1* (DOPyS <3) and *Group 2* (DOPyS ≥3)[§§§§]. Differences between groups calculated with a Mann-Whitney U test for non-parametric data

b) RoTA score

A statistically significant difference was demonstrated between the 2 groups when comparing the *RoTA score*. The overall median *RoTA* was 0.85 (0.40-1.28) for *Group 1* compared to 0.30 (0.20-0.50) for *Group 2* (p=0.04) (**Table 50**). When *RoTA* was plotted against individual DOPyS scores for all participants a clear trend was noted (**Figure 55**). There was no difference found in *RoTA* between those scoring ≤1 on the DOPyS and those scoring ≥3

[§§§§] Full data set can be found in **Appendix 5**

(p=0.85). However, there was a sharp rise in *RoTA* demonstrated for those scoring between >1 and <3 on the DOPyS. This group represents candidates with some experience of performing "real-life" polypectomy but who are not yet fully competent. There was no statistical difference found in *RoTA* between this group and the low scorers (DOPyS ≤1, p=0.07). However, there was a statistical difference demonstrated between this group and the highest scorers (DOPyS ≥3, p=<0.01).

c) Inter-rater reliability

There was a strong inter-rater correlation between the DOPyS assessors for overall competency at polypectomy (ρ = 0.76) and a very strong inter-rater correlation between the hand analysis assessors (**Figures 56 and 57**) for all parameters including; number of times clockwise rotation was applied (r=0.97), number of times anticlockwise rotation was applied (r= 0.98), number of times up or down angulation was used (r= 0.95), number of times right or left angulation was used (r= 0.95), total number of angulation movements (up/down and left-/right) (r= 0.96) and the total number of rotational movements (clockwise and anticlockwise) (r= 0.98).

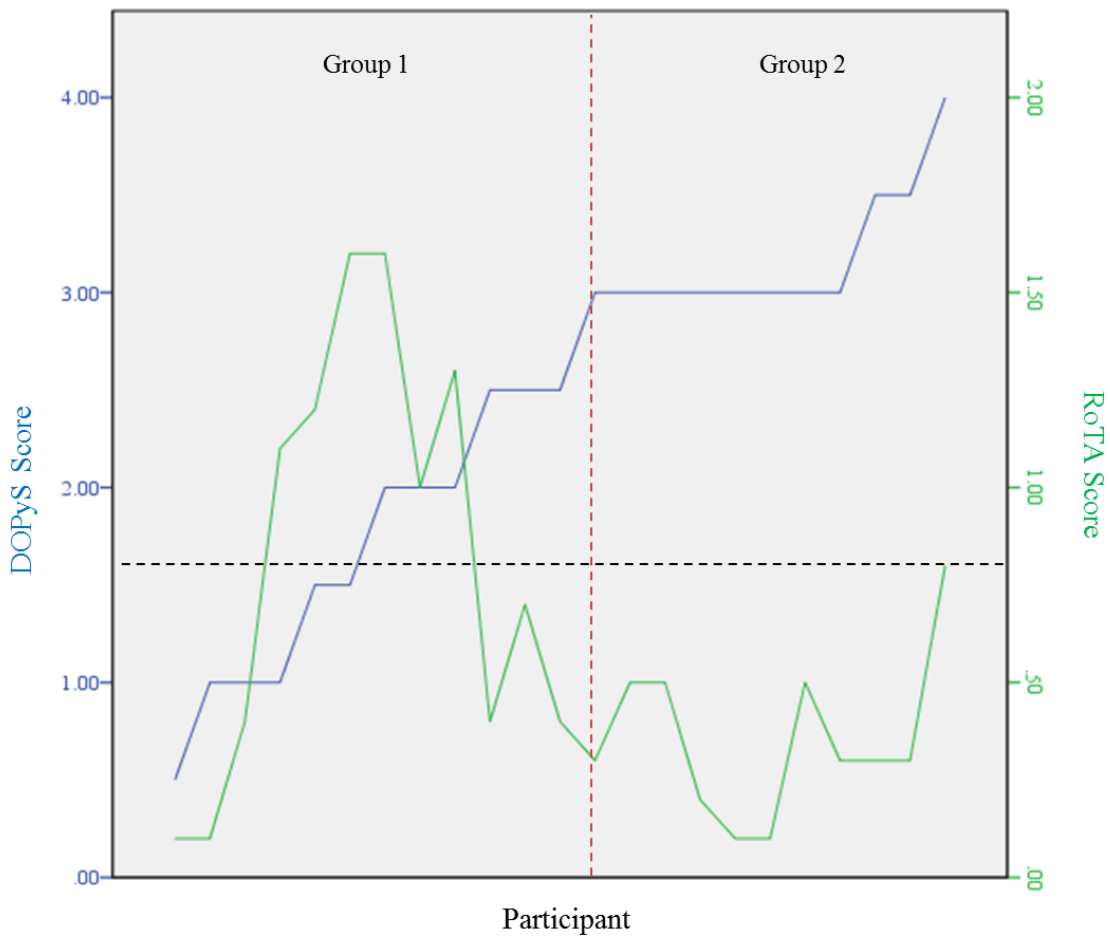


Figure 54: Comparison of RoTA score (green line) versus DOPyS assessment (blue line) for each participant. Red dotted line demonstrates the division between *Group 1* (DOPyS ≥3) and *2 (*DOPyS <3). Black dotted line RoTA score which equates to a DOPyS score of ≥3

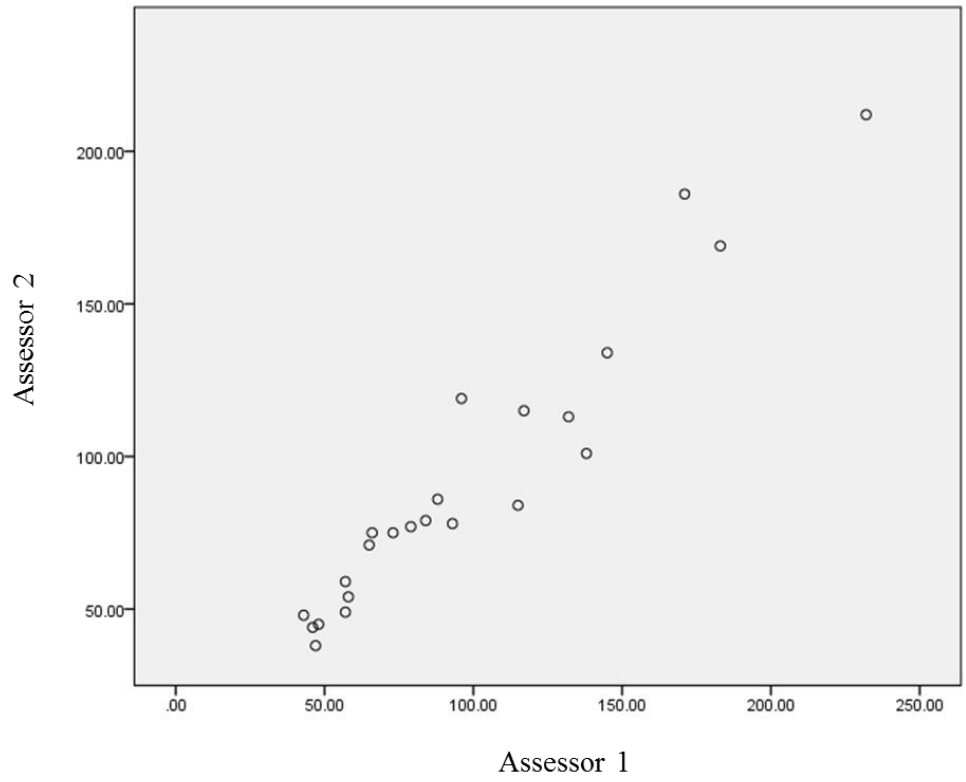


Figure 55: Assessor 1 vs 2 hand analysis scores for total number of tip movements (up/down and left/right) demonstrating good correlation between scoring

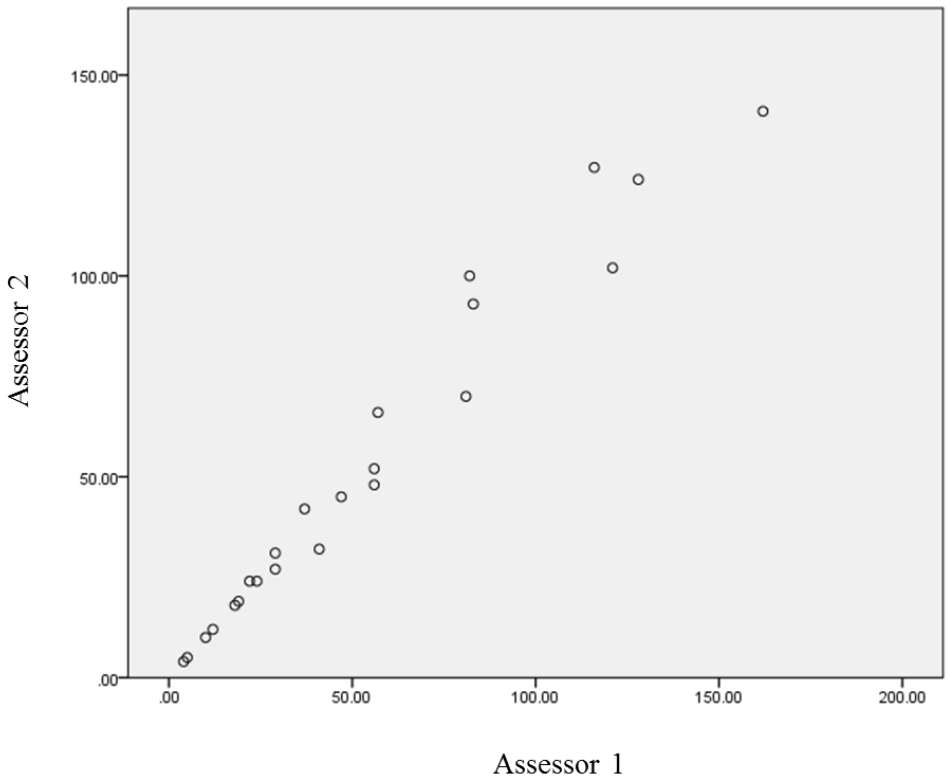


Figure 56: Assessor 1 vs 2 hand analysis scores for total number of rotational movements (Clockwise and Anti-clockwise) demonstrating good correlation between scoring

8.4 Discussion

a) Summary of results

The DOPyS is the current UK gold standard of subjective assessment used during colonic polypectomy (Gupta et al., 2011). This is the first study to describe a potential, objective assessment parameter. The WIMAT colonoscopy suitcase simulator was used to evaluate the differences in hand movements performed during colonic polypectomy. This identified that the *RoTA score* can accurately distinguish between candidates who obtain a competent DOPyS score (≥3) and those who are not yet DOPyS competent (<3). Our results also demonstrate a specific trend in the *RoTA score* that occurs with increasing DOPyS scores (**Figure 54**).

b) Difference in the hand movements used between novices and experts

There was no statistical difference demonstrated between *Group 1* and *Group 2* for any of the basic hand movement counts. It is likely that a difference may have been demonstrated if all candidates took the same amount of time to perform the task. However, in our study, *Group 2* completed the task in almost half the time that it took for *Group 1* to finish (p=0.01). The *RoTA score* of each group was therefore calculated in order to further examine the relationship between these hand movements and account for this time difference.

c) Experts use torque steering methods more than novices

As a result, a statistically significant difference in *RoTA score* was found between those scoring <3 and ≥3 on the DOPyS assessment. The *RoTA score* of DOPyS competent

candidates was lower than those who were not yet competent. When looking at individual candidates DOPyS scores against their *RoTA score*, an interesting trend was noted. Low and High DOPyS scorers produced similar *RoTA scores* which may have been due to the fact that low scorers were very inexperience in performing any type of endoscopy and may not have been aware of how to use rotational movements or torque steering. The intermediate DOPyS scorers may be using too much rotation in relation to tip control. This could reflect that they have experience of torque steering in scope navigation, but have yet to learn the correct balance of rotation to angulation need for more complex therapeutic interventions.

d) Inter-rater reliability

A key strength of this study is that we found good inter-rater correlation for both DOPyS scoring and hand analysis. This would suggest that the results of each assessment are reliable and an accurate reflection of performance. The set-up of the polypectomy simulation was identical for each participant in terms of polyp size, position and distance at which the polyp was inserted. This ensured that the only variable that was different each time was the individual performing the task.

e) Limitations of the study

One limitation of this study is that only one type of polyp task was examined. Although this was helpful to standardise each procedure, it may be the case that the *RoTA score* changes with different types of polyps. Polyps placed in differing positions (where either more or less torque is needed to perform the polypectomy) may change the relationship of *RoTA score* and DOPyS score. Another area of interest may be to examine the *RoTA score* at various stages of the procedure including; finding the polyp, positioning the scope, performing the

polypectomy and retrieval. Future work should identify if the *RoTA score* changes during these stages. A further limitation may have been that a short length of colon was used which required little expertise to navigate. Differences in *RoTA* may be further emphasized on a longer length of convoluted bowel or in the "real-life" situation where the polyp is in the transverse or ascending colon.

If *RoTA* scoring is to become a future objective performance indicator for colonic polypectomy, further work is needed to improve the way in which this data is collected. During this study we video recorded and manually counted hand movements. However, a more feasible approach may be to incorporate technologies such as kinematics and motion tracking. *RoTA* may even be an important performance parameter to incorporate into Virtual Reality endoscopic simulator software.

f) Conclusion

- There are currently no validated objective assessment measures for use in colonoscopic polypectomy
- There was no significant difference in the total number of hand movements performed between 2 cohorts of differing expertise
- The RoTA score can accurately distinguish between candidates who obtain a competent DOPyS score (≥3) and those who are not yet DOPyS competent (<3)
- Further work is required to determine the accuracy of RoTA in different types of polypectomy with different degrees of difficulty

Summary and future work

Summary

Colonoscopic polypectomy requires considerable practice and time to master; but, traditionally it has been taught at the expense of patient comfort and safety (Sedlack, 2005). Simulation may address this issue, but until now, there has been no validated model available for training. This thesis details the validation of a novel, ex-vivo simulator for polypectomy skills training. Furthermore, it presents preliminary evidence for ways in which this simulator can be used in the assessment of polypectomy skills. There were several aims and hypotheses that were constructed at the commencement of this work. We have been able to provide evidence to address these and have developed future strategies to identify areas of work which require further investigation.

Conclusion of aims

a) To develop a novel ex-vivo porcine simulator for colonoscopic polypectomy training

The first aim of this thesis was to develop a novel ex-vivo porcine simulator for colonoscopic polypectomy training. The WIMAT colonoscopy suitcase was developed in order to teach the skills needed to perform colonoscopic polypectomy. *Chapter 3* of this thesis demonstrates how the model has been adapted to focus on pedunculated polypectomy. Two levels of difficulty have been established in order to improve this model's applicability for a range of endoscopic expertise.

b) To test the content, construct and criterion validity of the simulator

The content validity of the WIMAT colonoscopy suitcase was evaluated using a previously validated realism survey (Sedlack et al., 2007b). Seventeen endoscopic experts scored the simulator favourably according to visual, anatomical, mechanical and overall realism. *Chapter 5* of this thesis details the construct and concurrent validation of this model. Eighty participants (20 novices, 20 intermediates, 20 advanced and 20 experts) completed a polypectomy task which involved excision of polyp A (simple) and polyp B (complex). These were retrospectively video scored using a shortened DOPyS assessment. Results showed a statistically significant difference between the DOPyS scores for polyp A versus B. In most cases, candidates scored higher for polyp A than B and completed the polyp A task in less time than polyp B. When DOPyS scores between the groups were compared, statistically significant differences were found in all cases apart from when the advanced group were compared with the experts. This therefore, provides good evidence for the construct validity of the WIMAT colonoscopy suitcase.

The concurrent validity of the suitcase was evaluated by comparing simulated DOPyS scores achieved on the model with "real-life" DOPyS score. This demonstrated that there was no statistical difference between "real-life" and simulated DOPyS scores for the majority of assessment parameters. This provides evidence for the concurrent validity of the simulator.

c) To evaluate the feasibility of measuring skills transfer using the simulator

Twenty participants were enrolled into a pilot study to evaluate the effect of polypectomy simulator training on clinical performance and the feasibility of conducting an RCT comparing the WIMAT colonoscopy suitcase with the GI mentor II (Simbionix, USA). Baseline DOPyS scores were compared to endpoint scores collected at 4 months post simulator training on either the ex-vivo or VR models. This demonstrated some degree of skills transfer but results were limited by the fact that some participants were unable to complete the endpoint DOPyS assessment. Further recruitment to replace these candidates was not feasible and therefore, alterations to the study protocol have been made for consideration before conducting a larger RCT.

d) To evaluate if self-assessment is an accurate when using the simulator

During the construct validation trial, candidates were also asked to complete a self-assessment DOPyS after completing the polypectomy task. These were then compared to the DOPyS scores awarded by the independent assessors. There was poor correlation between self-assesssment and independent examiners scores. Novices tended to underestimate whereas experts tended to overestimate performance.

e) To establish a quantitative scoring system for colonic polypectomy using the simulator

During the construct validation trial, the hand movements used when performing a simulated polypectomy were video recorded and analysed for 23 participants. The were no statistically

significant differences between the total number of rotational and total number of angulation movements performed between those that scored well on the standard DOPyS assessment (*Group 2*) and those that did not (*Group 1*). However, *Group 2* took approximately half the time to complete the task compared to group 1 ($p=<0.01$) and therefore, the ratio of rotational movements to angulation movements was analysed, the *RoTA score*. There was a significant difference in *RoTA score* between the two groups. This preliminary work requires development in order to evaluate the *RoTA* at differing stages of the polypectomy process and in polyps with varying degrees of difficulty.

Conclusion of hypotheses

a) The simulator is valid tool for colonic polypectomy skills training

The WIMAT colonoscopy suitcase is a valid colonoscopic polypectomy trainer. The evidence in this thesis supports this hypothesis for several different modalities of validation including; content, construct and concurrent validity.

b) Measuring skills transfer to the clinical setting following simulator training is feasible

This thesis confirms that measuring skills transfer following simulator training with the WIMAT colonoscopy suitcase is feasible. The feasibility study presented in this thesis has highlighted the need to improve certain areas of the trial protocol before conducted a larger skills transfer trial.

c) Trainees can use the simulator to accurately self-assess their performance

This thesis has presented evidence that questions the accuracy of self-assessment when using the WIMAT colonoscopy suitcase. The poor correlation between expert and self-assessment for simulated colonoscopy may be improved by the use of video play-back analysis.

d) The simulator can be used to establish a quantitative scoring system

This thesis confirms that the simulator can be used to establish a quantitative scoring system. RoTA score needs further development in order to establish its validity and reliability across a range of simulated polypectomy tasks. In turn, this needs to evaluated in the "real-life" clinical setting.

Future work

a) Adapting the WIMAT colonoscopy suitcase to incorporate complex polypectomy

The first stage of adapting the simulator has recently been explored. Transanal endoscopic microsurgery (TEM) has emerged as a minimally invasive means of resecting rectal tumours (Saclarides et al., 1992). TEM is performed with the use of a complex operative system, which allows the application of all surgical techniques for dissection and suturing inside the rectal cavity up to a height of 20 cm from the dentate line (Buess, 1993). The polyp construction technique used in the WIMAT colonoscopy suitcase has been adapted to create sessile rectal lesion in conjunction the Transanal Endoscopic Operation (TEO®) equipment (Karl Storz, UK) (**Figure 58**). The next stage is to validate this simulator.

Further adaptations to the WIMAT colonoscopy suitcase will also focus on creating more realistic polyp lesions in order to practise the skills needed to perform EMR and ESD. Several groups have already begun to design this novel simulation (Wang et al., 2011, Chen et al., 2012). Wang *et al* has created pseudo polyps using an oesophageal variceal ligation device to simulate a protruding (0-Ip) lesion, and the pseudo polyp was transected with a snare cautery to simulate a depressed (0-IIc) lesion (Wang et al., 2011). Similarly, Yoshida *et al* have begun development of an ex-vivo bovine animal for colonic ESD which has the capacity for recreating blood flow to allow haemostasis to be practised (Yoshida et al., 2013). These advances will be utilised to enhance the WIMAT colonoscopy suitcase in future work.

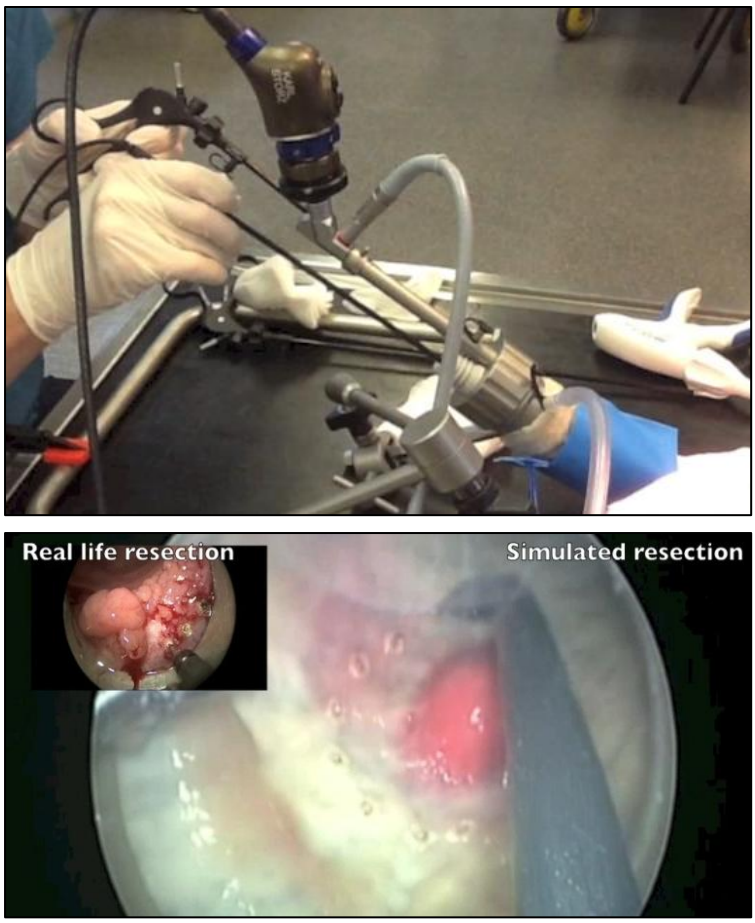


Figure 58: WIMAT TEO model, adapted from WIMAT colonoscopy suitcase model. Left = external view, Right = luminal view of simulated resection compared with real-life resection

b) Further validation of the WIMAT colonoscopy suitcase

Conducting a skills transfer study using this simulation model is feasible. Future work will concentrate on constructing a multi-centre RCT based on pilot work conducted during in this thesis.

c) Developing the RoTA assessment tool

The *RoTA score* assessment tool is currently in its infancy. The next stage is to assess if the results of the *RoTA score* are repeatable when evaluating polypectomy tasks of differing levels of complexity.

References

Abakken, L., Adamsen, S. & Kruse, A. 2000. Performance of a colonoscopy simulator: experience from a hands-on endoscopy course. *Endoscopy,* 32**,** 911-3.

ACT. 2012. *Cruelty to Animals Act (1876)* [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1872363/ [Accessed 31/7/2012 2012].

Ahlberg G., Hultcrantz, R., Jaramillo, E., Lindblom, A. & Arvidsson, D. 2005. Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? *Endoscopy,* 37**,** 1198-204.

Ahmad, N. A., Kochman, M. L., Long, W. B., Furth, E. E. & Ginsberg, G. G. 2002. Efficacy, safety, and clinical outcomes of endoscopic mucosal resection: a study of 101 cases. *Gastrointest Endosc,* 55**,** 390-6.

Almansa, C., Shahid, M. W., Heckman, M. G., Preissler, S. & Wallace, M. B. 2011. Association between visual gaze patterns and adenoma detection rate during colonoscopy: a preliminary investigation. *Am J Gastroenterol,* 106**,** 1070-4.

Atkin , W. S., Morson, B. C. & Cuzick, J. 1992. Long-term risk of colorectal cancer after excision of rectosigmoid adenomas. *N Engl J Med,* 326**,** 658-62.

Atkin, W. S. & Saunders, B. P. 2002. Surveillance guidelines after removal of colorectal adenomatous polyps. *Gut,* 51 Suppl 5**,** V6-9.

Ball, K., Lyons, N., Conlon, M. & Critchley, R. 2011. Strengthening medical appraisal: reactions to proposals for a new appraisal system to support revalidation. *Educ Prim Care,* 22**,** 409-14.

Bourke, M. 2011. Endoscopic mucosal resection in the colon: A practical guide. *Techniques in Gastrointestinal Endoscopy,* 13**,** 35-49.

Bowles, CJ., Leicester, R., Romaya, C., Swarbrick, E., Williams, CB., Epstein, O. 2004. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening today. *Gut,* 53, 277-82.

Brewster, L. P., Risucci, D. A., Joehl, R. J., Littooy, F. N., Temeck, B. K., Blair, P. G. & Sachdeva, A. K. 2008. Comparison of resident self-assessments with trained faculty and standardized patient assessments of clinical and technical skills in a structured educational module. *Am J Surg,* 195**,** 1-4.

Buess, G. 1993. Review: Transanal Endoscopic Microsurgery (TEM). *J R Coll Surg Edinb,* 38**,** 239-45.

Bujanda, L., Cosme, A., Gil, I. & Arenas-Mirave, J. I. 2010. Malignant colorectal polyps. *World J Gastroenterol,* 16**,** 3103-11.

Canada, C. 2013. *Treatments and Side Effects* [Online]. Available: http://www.google.co.uk/imgres?q=snare+polypectomy [Accessed 08/04/2013].

Cass, O. W., Freeman, M. L., Peine, C. J., Zera, R. T. & Onstad, G. R. 1993. Objective evaluation of endoscopy skills during training. *Ann Intern Med,* 118**,** 40-4.

Chen, M. J., Liu, C. Y., Chen, C. J., Shih, S. C. & Wang, H. Y. 2012. Simulating target lesion for endoscopic submucosal dissection training in a live pig model. *Endoscopy,* 44 Suppl 2 UCTN**,** E300-1.

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A. & Yang, G. Z. 2012. Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surg Endosc,* 26**,** 2003-9.

Clark, J., Orihuela-Espina, F., Sodergren, M., James, D. R., Karimyan, V., Teare, J., Darzi, A. & Yang, G. Z. 2013. A quantitative scale to define endoscopic torque control during natural orifice surgery. *Minim Invasive Ther Allied Technol,* 22**,** 17-25.

Classen, M. & Ruppin, H. 1974. Practical Endoscopy Training Using a New Gastrointestinal Phantom. *Endoscopy,* 6**,** 127-131.

Cohen, J., Cohen, S. A., Vora, K. C., Xue, X., Burdick, J. S., Bank, S., Bini, E. J., Bodenheimer, H., Cerulli, M., Gerdes, H., Greenwald, D., Gress, F., Grosman, I., Hawes, R., Mullin, G., Schnoll-Sussman, F., Starpoli, A., Stevens, P., Tenner, S. & Villanueva, G. 2006. Multicenter, randomized, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. *Gastrointest Endosc,* 64**,** 361-8.

Dafnis, G., Ekbom, A., Pahlman, L. & Blomqvist, P. 2001. Complications of diagnostic and therapeutic colonoscopy within a defined population in Sweden. *Gastrointest Endosc,* 54**,** 302-9.

Datta, V., Mandalia, M., Mackay, S. & Darzi, A. 2002. The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. *Surg Endosc,* 16**,** 1459-63.

Davis, D. A., Mazmanian, P. E., Fordis, M., Van Harrison, R., Thorpe, K. E. & Perrier, L. 2006. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA,* 296**,** 1094-102.

Desilets, D. J., Banerjee, S., Barth, B. A., Kaul, V., Kethu, S. R., Pedrosa, M. C., Pfau, P. R., Tokar, J. L., Varadarajulu, S., Wang, A., Wong Kee Song, L. M. & Rodriguez, S. A. 2011. Endoscopic simulators. *Gastrointestinal endoscopy,* 73**,** 861-7.

Duthie G., Drew P., Hughes M., Farouk R., Hodson R., Wedgwood K., Monson J. 1998. A UK training programme for nurse practitioner flexible sigmoidoscopy and a prospective evaluation of the practice of the first UK trained nurse flexible sigmoidoscopist. Gut, 43, 711-714.

Endodex. 2013. *Measuring quality in colonoscopy* [Online]. Available: http://www.d1261939-24233.cp.blacknight.com/ [Accessed 29/05/2013 2013].

Endoscopy. 2000. *Guidelines for the Training, Appraisal and Assessment of Trainees in GI Endoscopy.* [Online]. Available: http://www.thejag.org.uk/guidelines.htm.

Erlangen. 2012. *Endo-trainer* [Online]. Available: http://www.endo-trainer.de/endotrainer/english/endotrainer.php [Accessed 31/7/2012 2012].

Evans, A. W., Leeson, R. M. & Newton-John, T. R. 2002. The influence of self-deception and impression management on surgeons' self-assessment scores. *Med Educ,* 36**,** 1095.

Evans, A. W., Leeson, R. M. & Petrie, A. 2007. Reliability of peer and self-assessment scores compared with trainers' scores following third molar surgery. *Med Educ,* 41**,** 866-72.

Falchikov NB. 1989. Self-assessment in higher education: a metaanalysis. *Rev Educ Res* 59**,** 395-430.

Farrell, JJ., Friedmann, LS. 2005. Review article: the management of lower gastrointestinal bleeding. *Alimentary Pharmacology and Therapeutics,* 21, 1281-1298.

Fayez, R., Feldman, L. S., Kaneva, P. & Fried, G. M. 2010. Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters. *Surg Endosc,* 24**,** 1060-5.

Fearon, E. R. & Vogelstein, B. 1990. A genetic model for colorectal tumorigenesis. *Cell,* 61**,** 759-67.

Felsher, J. J., Olesevich, M., Farres, H., Rosen, M., Fanning, A., Dunkin, B. J. & Marks, J. M. 2005. Validation of a flexible endoscopy simulator. *Am J Surg,* 189**,** 497-500.

Fitzgerald, J. T., Gruppen, L. D. & White, C. B. 2000. The influence of task formats on the accuracy of medical students' self-assessments. *Acad Med,* 75**,** 737-41.

Fyock, C. J. & Draganov, P. V. 2010. Colonoscopic polypectomy and associated techniques. *World J Gastroenterol,* 16**,** 3630-7.

Gallegos-Orozco, J. F. & Gurudu, S. R. 2010. Complex colon polypectomy. *Gastroenterol Hepatol (N Y),* 6**,** 375-82.

Gastrotraining. 2013. *Basics of Colonoscopy* [Online]. Available: http://gastrotraining.com/category/gi-investigations [Accessed 20/05/2013 2013].

Gavin, DR., Valori, RM., Anderson, JT., Donnelly, MT., Williams, JG., Swarbrick, ET. 2013. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. *Gut,* 62, 242-9.

Gerson, L. B. & Van Dam, J. 2003. A prospective randomized trial comparing a virtual reality simulator to bedside teaching for training in sigmoidoscopy. *Endoscopy*, 35**,** 569-75.

Gilbert, D. A., Dimarino, A. J., Jensen, D. M., Katon, R., Kimmey, M. B., Laine, L. A., Macfadyen, B. V., Michaletz-Onody, P. A. & Zuckerman, G. 1992. Status evaluation: hot biopsy forceps. American Society for Gastrointestinal Endoscopy. Technology Assessment Committee. *Gastrointest Endosc,* 38**,** 753-6.

Gordon, M. J. 1991. A review of the validity and accuracy of self-assessments in health professions training. *Acad Med,* 66**,** 762-9.

Grantcharov, T. P., Carstensen, L. & Schulze, S. 2005. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *JSLS,* 9**,** 130-3.

Group, T. C. 2011. *General and Gastrointestinal: lower GI* [Online]. Available: http://www.thecgroup.com/products/lower-gi [Accessed 29/05/2013 2013].

Gupta, S., Anderson, J., Bhandari, P., Mckaig, B., Rupert, P., Rembacken, B., Riley, S., Rutter, M., Valori, R., Vance, M., Van Der Vleuten, C. P., Saunders, B. P. & Thomas-Gibson, S. 2011. Development and validation of a novel method for assessing competency in polypectomy: direct observation of polypectomy skills. *Gastrointest Endosc,* 73**,** 1232-9 e2.

Gupta, S., Bassett, P., Man, R., Suzuki, N., Vance, M. E. & Thomas-Gibson, S. 2012. Validation of a novel method for assessing competency in polypectomy. *Gastrointest Endosc,* 75**,** 568-75.

Gupta, S., Miskovic, D., Bhandari, P., Dolwani, S., Mckaig, B., Pullan, R., Rembacken, B., Riley, S., Rutter, MD., Suzuki, N., Tsiamoulos, Z., Valori, R., Vance, ME., Faiz, OD., Saunders, BP., Thomas-Gibson, S. 2013. A novel method for determining the difficulty of colonoscopic polypectomy. *Frontline Gastroent,* 4, 244-248.

Haggitt, R. C., Glotzbach, R. E., Soffer, E. E. & Wruble, L. D. 1985. Prognostic factors in colorectal carcinomas arising in adenomas: implications for lesions removed by endoscopic polypectomy. *Gastroenterology,* 89**,** 328-36.

Harrington, J. P., Murnaghan, J. J. & Regehr, G. 1997. Applying a relative ranking model to the self-assessment of extended performances. *Adv Health Sci Educ Theory Pract,* 2**,** 17-25.

Hassan, I., Osei-Agymang, T., Radu, D., Gerdes, B., Rothmund, M. & Fernandez, E. D. 2008. Simulation of laparoscopic surgery--four years' experience at the Department of Surgery of the University Hospital Marburg. *Wien Klin Wochenschr,* 120**,** 70-6.

Haycock, A., Koch, A. D., Familiari, P., Van Delft, F., Dekker, E., Petruzziello, L., Haringsma, J. & Thomas-Gibson, S. 2010. Training and transfer of colonoscopy skills: a multinational, randomized, blinded, controlled trial of simulator versus bedside training. *Gastrointest Endosc,* 71**,** 298-307.

Haycock, A. V., Bassett, P., Bladen, J. & Thomas-Gibson, S. 2009a. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. *Endoscopy,* 41**,** 952-8.

Haycock, A. V., Youd, P., Bassett, P., Saunders, B. P., Tekkis, P. & Thomas-Gibson, S. 2009b. Simulator training improves practical skills in therapeutic GI endoscopy: results from a randomized, blinded, controlled study. *Gastrointest Endosc,* 70**,** 835-45.

Health Commission. 2007. State of healthcare 2007: Improvements and challenges in services in England and Wales. [Online]. Available: http://www.wales.nhs.uk/documents/State_of_Healthcare_2007.pdf [Accessed 22/07/2014].

Hewitson, P., Glasziou, P., Irwig, L., Towler, B. & Watson, E. 2007. Screening for colorectal cancer using the faecal occult blood test, Hemoccult. *Cochrane Database Syst Rev***,** CD001216.

Higaki, S., Hashimoto, S., Harada, K., Nohara, H., Saito, Y., Gondo, T. & Okita, K. 2003. Long-term follow-up of large flat colorectal tumors resected endoscopically. *Endoscopy,* 35**,** 845-9.

Hochberger, J., Matthes, K., Maiss, J., Koebnick, C., Hahn, E. G. & Cohen, J. 2005. Training with the compact EASIE biologic endoscopy simulator significantly improves hemostatic technical skill of gastroenterology fellows: a randomized controlled comparison with clinical endoscopy training alone. *Gastrointestinal endoscopy,* 61**,** 204-15.

Hochberger, J., Neumann, M., Hohenberger, W. & Hahn, E. G. 1997. EASIE-Erlangen Education Simulation Model for Interventional Endoscopy--a new bio-training model for surgical endoscopy. *Biomedizinische Technik. Biomedical engineering,* 42 Suppl**,** 334.

Hong, S, P. 2012. How do I manage post-polypectomy bleeding? *Clin Endosc,* 45, 282-284.

Issenberg, S. B., Mcgaghie, W. C., Hart, I. R., Mayer, J. W., Felner, J. M., Petrusa, E. R., Waugh, R. A., Brown, D. D., Safford, R. R., Gessner, I. H., Gordon, D. L. & Ewy, G. A. 1999. Simulation technology for health care professional skills training and assessment. *JAMA,* 282**,** 861-6.

Jag. 2011. *Jag Trainee Certification Process – Colonoscopy* [Online]. Available: http://www.thejag.org.uk/LinkClick.aspx?fileticket=HUIBs4ajYEw%3d&tabid=220 [Accessed 02/04/2013 2013].

JAG. 2012. *Joint Advisory Group on Gastrointestinal Endoscopy* [Online]. Available: http://www.thejag.org.uk/TrainingforEndoscopists.aspx [Accessed 1/8/12 2012].

Kantsevoy, S. V., Adler, D. G., Conway, J. D., Diehl, D. L., Farraye, F. A., Kwon, R., Mamula, P., Rodriguez, S., Shah, R. J., Wong Kee Song, L. M. & Tierney, W. M. 2008. Endoscopic mucosal resection and endoscopic submucosal dissection. *Gastrointest Endosc,* 68**,** 11-8.

Kato, M., Gromski, M., Jung, Y., Chuttani, R. & Matthes, K. 2013. The learning curve for endoscopic submucosal dissection in an established experimental setting. *Surg Endosc,* 27**,** 154-61.

Kegel-Flom, P. 1975. Predicting supervisor, peer, and self ratings of intern performance. *J Med Educ,* 50**,** 812-5.

Kim, W. S., Woo, H. S., Ahn, W., Lee, K., Cho, J. H., Lee, D. Y. & Yi, S. Y. 2007. Non-clinical evaluation of the KAIST-Ewha Colonoscopy Simulator II. *Studies in health technology and informatics,* 125**,** 214-6.

Koch, A. D., Buzink, S. N., Heemskerk, J., Botden, S. M., Veenendaal, R., Jakimowicz, J. J. & Schoon, E. J. 2008a. Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. *Surg Endosc,* 22**,** 158-62.

Koch, A. D., Haringsma, J., Schoon, E. J., De Man, R. A. & Kuipers, E. J. 2008b. A second-generation virtual reality simulator for colonoscopy: validation and initial experience. *Endoscopy,* 40**,** 735-8.

Koch, A. D., Haringsma, J., Schoon, E. J., De Man, R. A. & Kuipers, E. J. 2012. Competence measurement during colonoscopy training: the use of self-assessment of performance measures. *Am J Gastroenterol,* 107**,** 971-5.

Koken. 2012. *Life simulation models* [Online]. Available: http://www.kokenmpc.co.jp/english/products/life_simulation_models/index.html [Accessed 31/7/2012 2012].

Kudo, S., Tamura, S., Nakajima, T., Yamano, H., Kusaka, H. & Watanabe, H. 1996. Diagnosis of colorectal tumorous lesions by magnifying endoscopy. *Gastrointest Endosc,* 44**,** 8-14.

Kudo, S. E. & Kashida, H. 2005. Flat and depressed lesions of the colorectum. *Clin Gastroenterol Hepatol,* 3**,** S33-6.

Kundel, H. L., Nodine, C. F., Conant, E. F. & Weinstein, S. P. 2007. Holistic component of image perception in mammogram interpretation: gaze-tracking study. *Radiology,* 242**,** 396-402.

Lee, S. H., Chung, I. K., Kim, S. J., Kim, J. O., Ko, B. M., Hwangbo, Y., Kim, W. H., Park, D. H., Lee, S. K., Park, C. H., Baek, I. H., Park, D. I., Park, S. J., Ji, J. S., Jang, B. I., Jeen, Y. T., Shin, J. E., Byeon, J. S., Eun, C. S. & Han, D. S. 2008. An adequate level of training for technical competence in screening and diagnostic colonoscopy: a prospective multicenter evaluation of the learning curve. *Gastrointest Endosc,* 67**,** 683-9.

Leslie, A., Carey, F. A., Pratt, N. R. & Steele, R. J. 2002. The colorectal adenoma-carcinoma sequence. *Br J Surg,* 89**,** 845-60.

Lieberman, D., Moravec, M., Holub, J., Michaels, L. & Eisen, G. 2008. Polyp size and advanced histology in patients undergoing colonoscopy screening: implications for CT colonography. *Gastroenterology,* 135**,** 1100-5.

Logan, R. F., Patnick, J., Nickerson, C., Coleman, L., Rutter, M. D. & Von Wagner, C. 2012. Outcomes of the Bowel Cancer Screening Programme (BCSP) in England after the first 1 million tests. *Gut,* 61**,** 1439-46.

Lohsiriwat, V. 2010. Colonoscopic Perforation: incidence, risk factors, management and outcome. *World J Gastroent,* 16, 425-430.

Long, V. & Kalloo, A. N. 2006. AccuTouch Endoscopy Simulator: development, applications and early experience. *Gastrointestinal endoscopy clinics of North America,* 16**,** 479-87.

Maagaard, M., Sorensen, J. L., Oestergaard, J., Dalsgaard, T., Grantcharov, T. P., Ottesen, B. S. & Larsen, C. R. 2011. Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. *Surg Endosc,* 25**,** 722-7.

Macdonald, J., Ketchum, J., Williams, R. G. & Rogers, L. Q. 2003. A lay person versus a trained endoscopist: can the preop endoscopy simulator detect a difference? *Surg Endosc,* 17**,** 896-8.

Mahmood, T. & Darzi, A. 2003. A study to validate the colonoscopy simulator. *Surg Endosc,* 17**,** 1583-9.

Maiss, J., Millermann, L., Heinemann, K., Naegel, A., Peters, A., Matthes, K., Hahn, E. G. & Hochberger, J. 2007. The compact EASIE is a feasible training model for endoscopic novices: a prospective randomised trial. *Dig Liver Dis,* 39**,** 70-8; discussion 79-80.

Martin, D., Regehr, G., Hodges, B. & McNaughton, N. 1998. Using videotaped benchmarks to improve the self-assessment ability of family practice residents. *Acad Med,* 73**,** 1201-6.

Maschuw, K., Hassan, I. & Bartsch, D. K. 2010. Surgical training using simulator. Virtual reality. *Chirurg,* 81**,** 19-24.

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

Maslow 1987. *Motivation and Personality,* New York, Harper and Row.

McConnell, R. A., Kim, S., Ahmad, N. A., Falk, G. W., Forde, K. A., Ginsberg, G. G., Jaffe, D. L., Makar, G. A., Long, W. B., Panganamamula, K. V. & Kochman, M. L. 2012. Poor discriminatory function for endoscopic skills on a computer-based simulator. *Gastrointest Endosc,* 76**,** 993-1002.

Mertz, A.J., Moss, A., Mcleod, D., Tran, K., Godfrey, C., Chandra, A., Bourke, M. J. 2013. A blinded comparison of the safety and efficacy of hot biopsy forceps electrocauterization and

conventional snare polypectomy for diminutive colonic polypectomy in a porcine model. *Gastrointest Endosc,* 77, 484-90.

Monkemuller, K., Neumann, H., Fry, L. C., Ivekovic, H. & Malfertheiner, P. 2008. Polypectomy techniques for difficult colon polyps. *Dig Dis,* 26**,** 342-6.

Monkemuller, K., Neumann, H., Malfertheiner, P. & Fry, L. C. 2009. Advanced colon polypectomy. *Clin Gastroenterol Hepatol,* 7**,** 641-52.

Moorthy, K., Munz, Y., Adams, S., Pandey, V. A., & Darzi, A. 2006. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *Am J Surg,* 192**,** 114-8.

Moorthy, K., Munz, Y., Orchard, T. R., Gould, S., Rockall, T. & Darzi, A. 2004. An innovative method for the assessment of skills in lower gastrointestinal endoscopy. *Surg Endosc,* 18**,** 1613-9.

Morris-Stiff, G. J., Sarasin, S., Edwards, P., Lewis, W. G. & Lewis, M. H. 2005. The European Working Time Directive: One for all and all for one? *Surgery,* 137**,** 293-7.

Morton, J. B. & Macbeth, W. A. 1977. Correlations between staff, peer and self assessments of fourth-year students in surgery. *Med Educ,* 11**,** 167-70.

Muto, T., Bussey, H. J. & Morson, B. C. 1975. The evolution of cancer of the colon and rectum. *Cancer,* 36**,** 2251-70.

Nelson, D. B., Bosco, J. J., Curtis, W. D., Faigel, D. O., Kelsey, P. B., Leung, J. W., Mills, M. R., Smith, P., Tarnasky, P. R., Vandam, J., Wang, K. K. & Wassef, W. Y. Technology status evaluation report: endoscopy simulators: May 1999. *Gastrointest Endosc. 2000 Jun;51(6):790-2.*

Neugut, A. I., Johnsen, C. M., Forde, K. A. & Treat, M. R. 1985. Recurrence rates for colorectal polyps. *Cancer,* 55**,** 1586-9.

O'brien, M. J., Winawer, S. J., Zauber, A. G., Gottlieb, L. S., Sternberg, S. S., Diaz, B., Dickersin, G. R., Ewing, S., Geller, S., Kasimian, D. 1990. The National Polyp Study. Patient and polyp characteristics associated with high-grade dysplasia in colorectal adenomas. *Gastroenterology,* 98**,** 371-9.

Obstein, K. L., Patil, V. D., Jayender, J., San Jose Estepar, R., Spofford, I. S., Lengyel, B. I., Vosburgh, K. G. & Thompson, C. C. 2011. Evaluation of colonoscopy technical skill levels by use of an objective kinematic-based system. *Gastrointest Endosc,* 73**,** 315-21, 321 e1.

Olympus. 2012. *Olympus Colonoscopy Simulator Community* [Online]. Available: http://www.colonoscopy-simulator.com/community/index.php [Accessed 01/08/2012 2012].

Pandey, V. A., Wolfe, J. H., Black, S. A., Cairols, M., Liapis, C. D. & Bergqvist, D. 2008. Self-assessment of technical skill in surgery: the need for expert feedback. *Ann R Coll Surg Engl,* 90**,** 286-90.

Papatheodoridis, G. V., Kapranos, N., Tzouvala, M., Triantafyllou, K., Elemenoglou, I. & Karamanolis, D. G. 1998. p53 and bcl-2 protein expression in rectosigmoid adenomas. *Am J Gastroenterol,* 93**,** 1136-40.

Physicians, R. C. O. 2012. *Joint Advisory Group on gastrointestinal endoscopy (JAG)* [Online]. [Accessed 1/8/2012].

Plooy, A. M., Hill, A., Horswill, M. S., Cresp, A. S., Watson, M. O., Ooi, S. Y., Riek, S., Wallis, G. M., Burgess-Limerick, R. & Hewett, D. G. 2012. Construct validation of a physical model colonoscopy simulator. *Gastrointest Endosc,* 76**,** 144-50.

Ponz De Leon, M. 1996. Genetic basis of tumour development. *Ital J Gastroenterol,* 28**,** 232-45.

Potter, J. D. 1999. Colorectal cancer: molecules and populations. *J Natl Cancer Inst,* 91**,** 916-32.

PRISMA. 2011. *PRISMA transparent reporting of systematic reviews and meta-analyses.* [Online]. Available: http://www.prisma-statement.org/statement.htm Accessed 25/08/2011].

Repici, A., Hassan, C., Vitetta, E., Ferrara, E., Manes, G., Gullotti, G., Princiotta, A., Dulbecco, P., Gaffuri, N., Bettoni, E., Pagano, N.,Rando, G., Strangio, G., Carlino, A., Romeo, F., de Paula, Pessoa, Ferreira D., Zullo, A., Ridola, L., Malesci, A. 2012. Safety of cold polypectomy for <10mm polyps at colonoscopy: a prospective multicentre study. *Endoscopy*, 44, 27-31.

Rex, D. K. 2010. Preventing colorectal cancer and cancer mortality with colonoscopy: what we know and what we don't know. *Endoscopy,* 42**,** 320-3.

Reynolds, T. & Kong, M. L. 2010. Shifting the learning curve. *BMJ,* 341**,** c6260.

Riley, S. A. 2008. *Colonoscopic Polypectomy And Endoscopic Mucosal Resection: A Practical Guide* [Online]. Available: http://www.bsg.org.uk/images/stories/docs/sections/endo/polypectomy_08.pdf [Accessed 02/04/2013 2013].

Rouilly, A. 2012. *Simulators, Anatomical Models And Charts For Clinical Skills And Training* [Online]. Available: http://www.adam-rouilly.co.uk/products.aspx?cid=365 [Accessed 31/7/2012 2012].

Saclarides, T. J., Smith, L., Ko, S. T., Orkin, B. & Buess, G. 1992. Transanal endoscopic microsurgery. *Dis Colon Rectum,* 35**,** 1183-91.

Saito, Y., Uraoka, T., Yamaguchi, Y., Hotta, K., Sakamoto, N., Ikematsu, H., Fukuzawa, M., Kobayashi, N., Nasu, J., Michida, T., Yoshida, S., Ikehara, H., Otake, Y., Nakajima, T., Matsuda, T. & Saito, D. 2010. A prospective, multicenter study of 1111 colorectal endoscopic submucosal dissections (with video). *Gastrointest Endosc,* 72**,** 1217-25.

Sakamoto, T., Saito, Y., Fukunaga, S., Nakajima, T. & Matsuda, T. 2011. Learning curve associated with colorectal endoscopic submucosal dissection for endoscopists experienced in gastric endoscopic submucosal dissection. *Dis Colon Rectum,* 54**,** 1307-12.

Schoenfield, P. 1999. Flexible Sigmoidoscopy by paramedical personnel. *J Clin Gastroenterol,* 28, 110-6.

Sedlack, R. E. 2005. Endoscopic simulation: where we have been and where we are going. *Gastrointest Endosc,* 61**,** 216-8.

Sedlack, R. E. 2011. Training to competency in colonoscopy: assessing and defining competency standards. *Gastrointestinal endoscopy,* 74**,** 355-366 e1-2.

Sedlack, R. E., Baron, T. H., Downing, S. M. & Schwartz, A. J. 2007a. Validation of a colonoscopy simulation model for skills assessment. *The American journal of gastroenterology,* 102**,** 64-74.

Sedlack, R. E., Baron, T. H., Downing, S. M. & Schwartz, A. J. 2007b. Validation of a colonoscopy simulation model for skills assessment. *Am J Gastroenterol,* 102**,** 64-74.

Sedlack, R. E. & Kolars, J. C. 2003. Validation of a computer-based colonoscopy simulator. *Gastrointest Endosc,* 57**,** 214-8.

Sedlack, R. E. & Kolars, J. C. 2004. Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. *Am J Gastroenterol,* 99**,** 33-7.

Sedlack, R. E., Kolars, J. C. & Alexander, J. A. 2004. Computer simulation training enhances patient comfort during endoscopy. *Clinical gastroenterology and hepatology,* 2**,** 348-52.

Simbionix. 2012. *GI mentor* [Online]. Available: http://simbionix.com/simulators/gi-bronch-gi-mentor/ [Accessed 01/08/2012 2012].

Singh, N., Harrison, M. & Rex, D. K. 2004. A survey of colonoscopic polypectomy practices among clinical gastroenterologists. *Gastrointest Endosc,* 60**,** 414-8.

Sivak, M. V. 2004. Polypectomy: looking back. *Gastrointest Endosc,* 60**,** 977-82.

Smith, S. G., Torkington, J., Brown, T. J., Taffinder, N. J. & Darzi, A. 2002. Motion analysis. *Surgical endoscopy,* 16**,** 640-5.

Source, G. 2013. *Incipiens polyp* [Online]. Available: http://www.gastrosource.com/maintenance?itemId=12688291&requestItemKey=12775645&tileDefinition=tiles.media_gallery_item&folderId=12775645 [Accessed 08/04/2013].

Stather, D. R., Maceachern, P., Rimmer, K., Hergott, C. A. & Tremblay, A. 2011. Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training. *Respirology,* 16**,** 698-704.

Statistics, O. F. N. 2011. *Colorectal Cancer, England, 2009* [Online]. Available: http://www.ons.gov.uk/ons/rel/cancer-unit/bowel-cancer-in-england/2009/sum-colorectal.html [Accessed 28/03/2013].

Stuart, M. R., Goldstein, H. S. & Snope, F. C. 1980. Self-evaluation by residents in family medicine. *J Fam Pract,* 10**,** 639-42.

Studiocode. 2013. *The Benchmark in Performance Improvement* [Online]. Available: http://www.studiocodegroup.com/wp-content/uploads/2010/02/MedSim_Website.pdf [Accessed 21/05/2013].

Sturm, L. P., Windsor, J. A., Cosman, P. H., Cregan, P., Hewett, P. J. & Maddern, G. J. 2008. A systematic review of skills transfer after surgical simulation training. *Ann Surg,* 248**,** 166-79.

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L. & Goldsmith, C. H. 2010. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol,* 10**,** 1.

Limbs and Things. 2012. *Skills training products for healthcare professionals* [Online]. Available: http://limbsandthings.com/ [Accessed 31/7/2012].

Tholoor, S., Tsagkournis, O., Basford, P., Bhandari, P. 2013. Managing difficult polyps: techniques and pitfalls. Available: http://www.annalsgastro.gr/index.php/annalsgastro/article/view/1253/1049 [Accessed 26/4/13].

Tolliver, K. A. & Rex, D. K. 2008. Colonoscopic polypectomy. *Gastroenterol Clin North Am,* 37**,** 229-51, ix.

Toshio, U., Fujii, T. 2008. Submucosal injection solution for gastrointestinal tract endoscopic mucosal resection and endoscopic submucosal dissection. Drug Des Devel Ther, 2, 131-138.

Tuggy, M. L. 1998. Virtual reality flexible sigmoidoscopy simulator training: impact on resident performance. *J Am Board Fam Pract,* 11**,** 426-33.

Van Nortwick, S. S., Lendvay, T. S., Jensen, A. R., Wright, A. S., Horvath, K. D. & Kim, S. 2010. Methodologies for establishing validity in surgical simulation studies. *Surgery,* 147**,** 622-30.

Van Putten, P. G., Ter Borg, F., Adang, R. P., Koornstra J. J., Romberg-Camps, M. J., Timmer, R., Poen, A. C., Kuipers, E. J., Van Leerdam, M.E. 2012. Nurse Endoscopists perform colonoscopies according to the international standards and with high patient satisfaction. *Endoscopy.* 44, 1127-32.

Van Stolk, R. U., Beck, G. J., Baron, J. A., Haile, R. & Summers, R. 1998. Adenoma characteristics at first colonoscopy as predictors of adenoma recurrence and characteristics at follow-up. The Polyp Prevention Study Group. *Gastroenterology,* 115**,** 13-18.

Voiosu, T., Ratiu, I., Hervoso, C, M., Furnari, M., Voiosu, A., Goldis, A., Mateescu, B., Tacconi, M., Dulbecco, P., Haider, A., Voiosu, M, R., Sporea, I. Polyp and endoscopist-related factors influencing complication rates in difficult colonic polypectomies: results of a multicentre observational study. *Rom J Intern Med,* 51, 172-8.

Vormbrock, K. & Monkemuller, K. 2012. Difficult colon polypectomy. *World J Gastrointest Endosc,* 4**,** 269-80.

Vosburgh, K. G., Stylopoulos, N., Estepar, R. S., Ellis, R. E., Samset, E. & Thompson, C. C. 2007. EUS with CT improves efficiency and structure identification over conventional EUS. *Gastrointest Endosc,* 65**,** 866-70.

Walsh, C. M., Sherlock, M. E., Ling, S. C. & Carnahan, H. 2012. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. *Cochrane database of systematic reviews,* 6**,** CD008237.

Wang, T. E., Wang, H. Y., Lin, C. C., Chen, T. Y., Chang, C. W., Chen, C. J. & Chen, M. J. 2011. Simulating a target lesion for endoscopic submucosal dissection training in an ex vivo pig model. *Gastrointest Endosc,* 74**,** 398-402.

Ward, M., Macrae, H., Schlachta, C., Mamazza, J., Poulin, E., Reznick, R. & Regehr, G. 2003. Resident self-assessment of operative performance. *Am J Surg,* 185**,** 521-4.

Waye, J. D. 2005. Advanced polypectomy. *Gastrointest Endosc Clin N Am,* 15**,** 733-56.

West, N. J., Poullis, A. P. & Leicester, R. J. 2008. The NHS Bowel Cancer Screening Programme - a realistic approach with additional benefits. *Colorectal Dis,* 10**,** 708-14.

Williams, C. B. 1991. Small polyps: the virtues and the dangers of hot biopsy. *Gastrointest Endosc,* 37**,** 394-5.

Wilson, M. R., Vine, S. J., Bright, E., Masters, R. S., Defriend, D. & McGrath, J. S. 2011. Gaze training enhances laparoscopic technical skill acquisition and multi-tasking performance: a randomized, controlled study. *Surg Endosc,* 25**,** 3731-9.

Winawer, S. J., O'brien, M. J., Waye, J. D., Kronborg, O., Bond, J., Fruhmorgen, P., Sobin, L. H., Burt, R., Zauber, A. & Morson, B. 1990. Risk and surveillance of individuals with colorectal polyps. WHO Collaborating Centre for the Prevention of Colorectal Cancer. *Bull World Health Organ,* 68**,** 789-95.

Winawer, S. J., Zauber, A. G., Ho, M. N., O'Brien, M. J., Gottlieb, L. S., Sternberg, S. S., Waye, J. D., Schapiro, M., Bond, J. H., Panish, J. F. 1993. Prevention of colorectal cancer by colonoscopic polypectomy. The National Polyp Study Workgroup. *N Engl J Med,* 329**,** 1977-81.

Woo, H. S., Kim, W. S., Ahn, W., Lee, D. Y. & Yi, S. Y. 2008a. Haptic interface of the KAIST-Ewha colonoscopy simulator II. *IEEE Trans Inf Technol Biomed,* 12**,** 746-53.

Woo, H. S., Kim, W. S., Ahn, W., Lee, D. Y. & Yi, S. Y. 2008b. Haptic interface of the KAIST-Ewha colonoscopy simulator II. *IEEE transactions on information technology in biomedicine,* 12**,** 746-53.

Workshop, P. I. T. P. 2003. The Paris endoscopic classification of superficial neoplastic lesions: esophagus, stomach, and colon: November 30 to December 1, 2002. *Gastrointest Endosc,* 58**,** S3-43.

Yoshida, N., Yagi, N., Inada, Y., Kugai, M., Kamada, K., Katada, K., Uchiyama, K., Ishikawa, T., Takagi, T., Handa, O., Konishi, H., Kokura, S., Inoue, K., Wakabayashi, N., Abe, Y., Yanagisawa, A. & Naito, Y. 2013. Possibility of ex vivo animal training model for colorectal endoscopic submucosal dissection. *Int J Colorectal Dis,* 28**,** 49-56.

Appendix

Appendix 1

A systematic review of validity testing in colonoscopic simulation

Domain	Keywords
Endoscopy	(“endoscopy” OR “endoscopic” OR “colonoscopy”, “colonoscopic” OR “diagnostic colonoscopy” OR “therapeutic colonoscopy” OR “polypectomy” OR “polyps” OR “adenoma”)
Simulation	**AND** (“simulation” OR “simulator” OR “virtual reality simulation” OR “ex-vivo simulation” OR “live animal simulation” OR “bench models” OR “training” OR “assessment”)
Validation	**AND** (“validation” OR “validity” OR “face validity” OR “expert validity” OR “content validity” OR “construct validity” OR “concurrent validity” OR “criterion validity” OR “predictive validity”)

Appendix 1: Search strategy for systematic review

Appendix 2

The WIMAT colonoscopy suitcase, testing its content validity

1. Demographics

Age ○ 20-24 25-30 ○ 31-35 ○ 36-40 ○ 41+ ○

Gender Male ○ Female ○ Prefer not to say ○

Training Level (or equivalent) FP1-2 ○ CT1-2 ○ ST3-4 ○ ST5-6+ ○ ST7 ○

Consultant ○ Nurse Specialist ○ Medical Student ○ Other ○

If other, please state:

If ST3-ST7, do you have a National Training Number Yes ○ No ○ Unsure ○

What is your specialty or planned future specialty?

2. Simulator Experience

Have you used this simulator before? Yes ○ No ○ Unsure ○

If yes, how many times? 1-2 ○ 3-4 ○ 5+ ○ Unsure ○

Have you used any other endoscopy simulators before, if so how many times?

○ Virtual Reality Model number of times

○ Animal tissue Model number of times

○ Cadaveric Model number of times

○ Live animal Model number of times

○ Plastic/ Bench Model number of times

○ Other number of times

3. Training

JAG Courses attended if trainee or taught on if trainer:	Foundation Endoscopy	Yes ○	No ○	Unsure ○
	Basic Skills in Colonoscopy	Yes ○	No ○	Unsure ○
	Intermediate skills	Yes ○	No ○	Unsure ○
	Advanced Skills	Yes ○	No ○	Unsure ○
	Polypectomy Course	Yes ○	No ○	Unsure ○

If a trainee, have you received formal training outside JAG courses? Yes ○ No ○ Unsure ○ NA ○

If, yes please give details

If a trainee, how many of the following procedures have you performed (if a trainer, please give an estimated figure in the box)?

Colonoscopy		**0-30** ○	**31-60** ○	**61-90** ○	**91-120** ○
OGD		**0-30** ○	**31-60** ○	**61-90** ○	**91-120** ○
Flexi Sigmoidoscopy		**0-30** ○	**31-60** ○	**61-90** ○	**91-120** ○
Colonoscopy		**121-150** ○	**151-180** ○	**181-210** ○	**211+** ○
OGD		**121-150** ○	**151-180** ○	**181-210** ○	**211+** ○
Flexi Sigmoidoscopy		**121-150** ○	**151-180** ○	**181-210** ○	**211+** ○
Polypectomy		**0-10** ○	**11-20** ○	**21-30** ○	**31-40** ○
Biopsies		**0-10** ○	**11-20** ○	**21-30** ○	**31-40** ○
EUS		**0-10** ○	**11-20** ○	**21-30** ○	**31-40** ○
ERCP		**0-10** ○	**11-20** ○	**21-30** ○	**31-40** ○
Polypectomy		**41-50** ○	**51-60** ○	**61-70** ○	**71+** ○
Biopsies		**41-50** ○	**51-60** ○	**61-70** ○	**71+** ○
EUS		**41-50** ○	**51-60** ○	**61-70** ○	**71+** ○
ERCP		**41-50** ○	**51-60** ○	**61-70** ○	**71+** ○

	Strongly Disagree			Neutral			Strongly Agree
Visual Realism							
1. The model's mucosa appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
2. The endoscopic view appears the same as during human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
3. The model's polyp appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
4. The model's bleeding appears realistic as compared to human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Anatomic Realism							
5. The model's haustral/ folds are the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
6. The model's pedunculated polyp is anatomically the same as in real life	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
7. The model's sessile polyp is anatomically the same as in real life	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Mechanical Realism							
8. Resistance to scope advancement feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
9. Paradoxical scope motion occurs as it does in human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
10. Control of the Snare feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
11. Handling the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
12. Diathermy of the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
13. Raising the mucosa around the polyp feels the same as human endoscopy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
Summary evaluation							
14. Overall I felt the ex vivo model to simulate human polypectomy with great accuracy	1 ☐	2 ☐	3 ☐	4 ☐	5 ☐	6 ☐	7 ☐
	Much easier			**Same**			**Much more difficult**
15. Compared to human polypectomy, the simulation's level of technical difficulty	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

I would be interested in receiving information about future simulation based course ☐

I would be interested in participating in simulation based research ☐

Email address: _____

Any comments: _____

_____

_____

Appendix 2a: Questionnaire and data collection platform used for the content validation study adapted from (Sedlack et al., 2007b)

Question		Participant																
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Visual	**1**	7	6	3	6	5	6	5	6	6	6	5	4	6	5	6	6	7
	2	5	6	3	5	5	6	5	6	5	5	5	5	6	5	4	6	7
	3	6	6	4	5	5	5	5	6	5	5	6	5	4	5	6	6	5
	4	5	6	4	5	4	6	7	6	6	5	6	1	5	4	5	6	5
Anatomical	**5**	4	6	4	5	5	6	5	4	3	5	5	4	5	3	4	4	5
	6	6	6	5	4	6	5	5	5	3	5	5	4	5	6	3	6	3
	7	5	6	5	3	4	3	5	5	3	5	6	4	4	4	4	5	4
Mechanical	**8**	3	6	4	4	4	5	5	2	5	4	6	1	6	4	4	6	4
	9	3	6	3	4	6	6	3	4	4	4	6	5	4	3	6	6	4
	10	6	7	6	4	6	7	7	6	5	5	6	4	6	6	5	6	7
	11	6	7	5	5	6	7	6	6	6	4	6	4	6	5	5	6	7
	12	6	7	6	4	4	1	5	6	5	4	6	4	6	5	5	4	7
	13	6	6	6	4	5	6	6	6	7	5	6	4	6	5	5	5	7
Summary	**14**	6	5	6	5	6	5	6	6	6	5	6	5	6	5	6	6	5
	15	3	6	5	3	3	3	4	4	2	4	3	3	4	4	4	6	4

Appendix 2b: Dataset from content validation study

Appendix 3:

Testing the construct and concurrent validity of the WIMAT colonoscopy suitcase

Part of the research infrastructure for Wales funded by the National Institute for Social Care and Health Research, Welsh Government.
Yn rhan o seilwaith ymchwil Cymru a ariannir gan y Sefydliad Cenedlaethol ar gyfer Ymchwil Gofal Cymdeithasol ac Iechyd, Llywodraeth Cymru

NISCHR
Gwasanaeth Moeseg Ymchwil | RES | Research Ethics Service

Research Ethics Committee (REC) for Wales
Sixth Floor, Churchill House
17 Churchill Way
Cardiff CF10 2TW
Telephone : 029 2037 6829
Fax : 029 2037 6824

E-mail : corinne.scott@wales.nhs.uk

Website : www.nres.nhs.uk

25 April 2013

Mr. Jared Torkington
Consultant Colorectal Surgeon
University Hospital of Wales
WIMAT
Cardiff Medicentre
Heath Park
Cardiff CF14 4UJ

Dear Mr. Torkington

Full title of project: **The WIMAT colonoscopy suitcase has construct and concurrent validity for colonoscopic polypectomy skills training: A prospective, cross sectional study.**

Thank you for seeking the Committee's advice about the above project.

You provided the following documents for consideration:

- Covering letter signed yourself
- Summary protocol, no version or date

I enclose a copy of our leaflet, "Defining Research", which explains how we differentiate research from other activities. This project would not considered to be research according to this guidance. Therefore it does not require ethical review by a NHS Research Ethics Committee.

You should check with the University Hospital of Wales what other review arrangements or sources of advice apply to projects of this type. Guidance may be available from the clinical governance office.

This letter should not be interpreted as giving a form of ethical approval or any endorsement of the project, but it may be provided to a journal or other body as evidence that ethical approval is not required under NHS research governance arrangements.

However, if you, your sponsor/funder or any NHS organisation feels that the project should be managed as research and/or that ethical review by a NHS REC is essential, please write setting out your reasons and we will be pleased to consider further.

Where NHS organisations have clarified that a project is not to be managed as research, the

Research Governance Framework states that it should not be presented as research within the NHS.

Yours sincerely

Dr Corinne Scott
Committee Co-ordinator

Enclosure: *NISCHR RES leaflet – "Defining Research"*

Appendix 3a: Research Ethics Committee for Wales ethical approval response letter

Participant ID .....

Assessor .....

Assigned group .....

Validity and Self-Assessment Survey

1. Demographics

Age ☐

Gender Male ○ Female ○

Training Level (or equivalent) FP1-2 ○ CT1-2 ○ ST3-4 ○ ST5-6 ○ ST7-8 ○

Consultant ○ Nurse Specialist ○ Medical Student ○ SAS ○ Other ○

If other, please state: ☐

If ST3-ST7, do you have a National Training Number Yes ○ No ○ NA ○

What is your specialty or planned future specialty? ☐

Are you an accredited bowel screening Colonoscopist Yes ○ No ○

2. Simulator Experience

Have you used this simulator before? Yes ○ No ○ Unsure ○

If yes, how many sessions? 1-2 ○ 3-4 ○ 5+ ○ Unsure ○

Have you used any other endoscopy Simulators before, if so how many times?

○	Virtual Reality Model	number of times ☐
○	Animal tissue Model	number of times ☐
○	Cadaveric Model	number of times ☐
○	Live animal Model	number of times ☐
○	Plastic/ Bench Model	number of times ☐
○	Other	number of times ☐

3. Training

JAG Courses attended if trainee or taught on if trainer:	Foundation Endoscopy	Yes ○	No ○	Unsure ○
	Basic Skills in Colonoscopy	Yes ○	No ○	Unsure ○
	Intermediate skills	Yes ○	No ○	Unsure ○
	Advanced Skills	Yes ○	No ○	Unsure ○
	Polypectomy Course	Yes ○	No ○	Unsure ○

Others:

How many of the following procedures have you performed? PLEASE GIVE EXACT NUMBERS IN THE SQUARE BOX IF YOU ARE ABLE.

Colonoscopy		0-50 ○	51-100 ○	101-150 ○	151-200 ○	200-500 ○	>500 ○
Flexible Sigmoidoscopy		0-50 ○	51-100 ○	101-150 ○	151-200 ○	200-500 ○	>500 ○
Sessile Polypectomy		0-20 ○	21-40 ○	41-60 ○	61-80 ○	81-100 ○	>100 ○
Pedunculated Polypectomy		0-20 ○	21-40 ○	41-60 ○	61-80 ○	81-100 ○	>100 ○

If you are a **Trainee**, have you previously completed a formative Direct Observation of Polypectomy Skills (DOPyS) form with your trainer in the last 3 months? Yes ○ No ○ Unsure ○

If **yes**, what is your average score over the last 3 months for stalked polypectomy for the following? (4=Highly skilled, 3=comptent and safe, 2=some standrds met, 1=standards not yet met, 0=unsure)

Achieving optimal polyp position	
Optimising your view of the polyp	
Accurately positioning snare over polyp head	
Placing the snare at an appropriate position on polyp stalk	
Trapping the appropriate amount of tissue in the snare	
Using the appropriate degree of diathermy to remove polyp	
Your overall competency at polypectomy	

4. Self-Assessment

HOW WOULD YOU ASSESS YOUR OWN ABILITY FOR THE FOLLOWING PARAMETERS WHEN PERFORMING A POLYPECTOMY?

Achieving optimal polyp position	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Optimising your view of the polyp	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Accurately positioning snare over polyp head	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Placing the snare at an appropriate position on polyp stalk	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Trapping the appropriate amount of tissue in the snare	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Using the appropriate degree of diathermy to remove polyp	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Retrieval of polyp	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐
Your overall competency at polypectomy	Highly skilled	☐
	Competent and safe	☐
	Some standards met	☐
	Standards not yet met	☐
	Unsure	☐

5. Other

I would be interested in receiving information about future simulation based course ☐

I would be interested in participating in simulation based research ☐

I would be willing to provide evidence of previous DOPyS Scores (if available) ☐

I would be willing to be contacted by email ☐

Email address: _____

Any comments:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Thank you for taking the time to participate in this survey

Appendix 3b: Questionnaire and data collection platform for the construct validation studies

DOPyS Descriptors-Generic

Scale 4: Highly Skilled Performance

1. Ensures good (5-11 o'clock axis) polyp position with no errors. Attempts made at position correction throughout the procedure.

2. Maintains clear polyp views throughout the procedure.

3. Determines the full extent of the lesion, using adjunctive measures where appropriate.

4. Uses most appropriate polypectomy technique safely with no errors.

5. Maintains stable scope position throughout polypectomy. This may involve asking an assistant to hold the scope in position to provide a stable platform for polypectomy.

6. Checks all polypectomy equipment is available and functioning with correct settings prior to the procedure.

7. Checks snare prior to introduction into the scope and ensures that snare is marked appropriately on the snare handle.

8. Maintains effective communication with the staff and addresses patient's concerns.

9. Checks diathermy settings are appropriate and ensures diathermy equipment is available and working. Ensures pad is attached to patient, foot pedal is accessible., no contraindication to diathermy.

10. Accurately photo documents pre and post polypectomy if appropriate

Scale 3: Competent and safe throughout procedure, no uncorrected errors

1. Maintains 5-11 o'clock axis during procedure with attempts at position correction.

2. Attempts to obtain clear polyp views through aspiration, insufflation and lens wash.

3. Determines the full extent of the lesion, may not use adjunctive measures.

4. Uses appropriate polypectomy technique safely based on size, site and morphology.

5. Adjusts and stabilises scope position prior to polypectomy.

6. Checks polypectomy equipment is available and functioning.

7. Checks snare prior to introduction into the scope and ensures handle is marked.

8. Maintains effective communication either with the staff or patient.

9. Checks diathermy settings are appropriate. Ensures diathermy equipment is available and working, Epad is attached to patient, foot pedal is accessible, no contraindication to diathermy.

10. Photo documents pre and post polypectomy if appropriate

Scale 2: Some standards not yet met, aspects to be improved, some errors uncorrected

1. Does not maintain 5-11 o'clock axis. Few attempts made at position correction.

2. Clear polyp views not maintained.

3. Does not determine or visualise full extent of the polyp or fails to recognise features suggestive of malignancy.

4. Chooses inappropriate polypectomy technique.

5. Scope not stabilised adequately. Little or no attempts made at use of adjunctive techniques.

6. Does not check essential polypectomy equipment is available and functioning prior to the procedure.

7. Does not check snare functioning and marking prior to introduction into the scope.

8. Fails to give clear instructions to endoscopy staff during the procedure or ignores patient concerns.

9. Does not check diathermy settings.

10. Does not photo document pre and post polypectomy where appropriate

Scale 1: Accepted standards not yet met, frequent errors uncorrected

1. Does not maintain polyp in the optimal position at any time during the procedure.

2. Poor polyp views throughout the procedure with no attempts at correction.
3. No attempts made at determining or visualising full extent of the polyp. Attempts polypectomy on lesions which are unlikely to be endoscopically resectable.

4.Inappropriate polypectomy technique. Uses inappropriate diathermy settings. Uses diathermy or hot biopsy technique unsafely or inappropriately.

5. Unstable scope position throughout procedure with no attempts made at correction.

6. Does not check for any polypectomy equipment

7. Does not check snare functioning and marking prior to introduction into the scope.

8. Does not communicate with the endoscopy staff or patient throughout the procedure.

9. Makes no attempt to check, or uses inappropriate diathermy settings.

10. Does not photo document when appropriate

Appendix 3c: DOPyS descriptors used by assessors during validation trials

PARTICIPANT	AGE	GENDER	SPECIALTY	LEVEL
N1	29	M	ENT	CT1-2
N2	29	M	ENT	CT1-2
N3	49	M	GEN	CT1-2
N4	24	M	SURGERY	FP2
N5	25	M	SURGERY	FP2
N6	25	M	SURGERY	FP2
N7	29	M	SURGERY	ST2
N8	28	M	ENT	CT2
N9	27	M	SURGERY	FP2
N10	28	M	SURGERY	CT1-2
N11	28	F	UNSURE	CT1-2
N12	32	M	ORTHO	CT1-2
N13	34	M	ORTHO	CT1-2
N14	27	M	ENT	CT1-2
N15	27	M	SURGERY	CT1-2
N16	28	F	COLORECTAL	OTHER
N17	36	M	TRANSPLANT	OTHER
N18	38	M	VASC	LAT
N19	29	F	SURGERY	CT1-2
N20	28	M	UPPER GI	CT1-2
I1	29	M	GASTRO	ST3-4
I2	42	M	SURGERY	ST7
I3	38	M	COLORECTAL	LAT
I4	27	M	UPPER GI	CRF
I5	28	M	UPPER GI	CRF
I6	30	M	COLORECTAL	ST3-4
I7	33	M	COLORECTAL	ST5-6
I8	35	M	UPPER GI	ST5-6
I9	31	F	COLORECTAL	ST5-6
I10	38	M	COLORECTAL	ST5-6

I11	31	M	UPPER GI	ST3-4
I12	38	M	COLORECTAL	ST5-6
I13	32	M	UPPER GI	ST5-6
I14	34	M	GASTRO	ST5-6
I15	31	M	COLORECTAL	ST3-4
I16	35	M	GASTRO	ST3-4
I17	34	F	GASTRO	ST5-6
I18	36	M	VASCULAR	ST7-8
I19	32	M	GASTRO	ST3-4
I20	34	M	GASTRO	ST5-6
A1	40	M	COLORECTAL	CON
A2	39	M	GASTRO	CON
A3	55	M	COLORECTAL	CON
A4	44	F	COLORECTAL	CON
A5	48	M	COLORECTAL	CON
A6	44	M	SURGERY	CON
A7	54	M	COLORECTAL	CON
A8	42	F	COLORECTAL	CON
A9	40	F	COLORECTAL	CON
A10	46	M	COLORECTAL	CON
A11	41	M	SURGERY	CON
A12	54	M	SURGERY	CON
A13	45	M	GASTRO	CON
A14	50	F	GASTRO	CON
A15	52	M	GASTRO	CON
A16	37	M	SURGERY	CON
A17	41	M	GASTRO	CON
A18	42	M	GASTRO	CON
A19	35	M	GASTRO	CON
A20	45	M	GASTRO	CON
E1	43	M	GASTRO	CON
E2	43	M	GASTRO	CON

E3	51	M	GASTRO	CON
E4	46	M	GASTRO	CON
E5	43	M	GASTRO	CON
E6	48	M	GASTRO	CON
E7	49	M	GASTRO	CON
E8	43	M	COLORECTAL	CON
E9	37	M	GASTRO	CON
E10	58	M	COLORECTAL	CON
E11	48	M	GASTRO	CON
E12	40	M	GASTRO	CON
E13	50	M	GASTRO	CON
E14	36	M	GASTRO	CON
E15	47	M	COLORECTAL	CON
E16	57	M	GASTRO	CON
E17	42	M	GASTRO	CON
E18	46	M	GASTRO	CON
E19	54	M	GASTRO	CON
E20	51	F	GASTRO	CON

Appendix 3d: Group demographic data for participants completing the construct validation study where, N = Novice, I = Intermediate, A = Advanced, E = Expert

PARTICIPANT	VR	ANIMAL	CADAVERIC	LIVE ANIMAL	BENCH	OTHER
N1	0	0	0	0	0	0
N2	0	0	0	0	0	0
N3	3	0	0	0	0	0
N4	0	0	0	0	0	0
N5	0	0	0	0	0	0
N6	0	0	0	0	0	0
N7	3	0	0	0	3	0
N8	0	0	0	0	0	0
N9	0	0	0	0	0	0
N10	0	0	0	0	0	0
N11	0	0	0	0	3	0
N12	0	0	0	0	0	0
N13	0	2	0	0	3	0
N14	0	0	0	0	0	0
N15	0	0	0	0	0	0
N16	0	0	0	0	0	0
N17	0	0	0	0	0	0
N18	0	0	0	0	0	0
N19	0	0	0	0	0	0
N20	0	0	0	0	0	0
I1	0	1	0	0	0	0
I2	0	0	0	0	0	0
I3	0	0	0	0	0	0
I4	0	3	0	0	3	0
I5	0	5	0	0	10	0
I6	5	0	0	0	0	0
I7	2	0	0	0	2	0
I8	1	0	0	2	0	0
I9	2	0	0	0	1	0
I10	1	4	0	0	1	0

I11	3	3	0	0	2	0
I12	10	0	0	0	0	0
I13	1	0	0	0	0	0
I14	2	0	0	0	0	0
I15	1	0	0	0	0	0
I16	1	1	0	0	0	0
I17	1	1	0	0	0	0
I18	1	0	0	0	2	0
I19	1	0	0	0	1	0
I20	3	0	0	0	0	0
A1	3	2	0	0	0	0
A2	2	0	0	0	6	0
A3	2	0	0	0	0	0
A4	0	0	0	0	0	0
A5	0	0	0	0	0	0
A6	0	0	0	0	0	0
A7	0	0	0	0	0	0
A8	2	4	0	0	2	0
A9	0	0	0	0	2	0
A10	0	0	0	0	2	0
A11	0	0	0	0	1	0
A12	5	0	0	3	0	0
A13	1	4	0	0	0	0
A14	1	5	0	0	0	0
A15	0	0	0	0	0	0
A16	3	0	0	0	3	0
A17	0	0	0	0	0	0
A18	0	1	0	0	0	0
A19	2	0	0	0	0	0
A20	3	4	0	2	6	0
E1	10	5	0	0	10	0
E2	50	0	0	0	150	0

E3	100	100	0	0	0	0
E4	100	10	0	0	100	0
E5	4	10	0	0	4	0
E6	5	5	0	0	0	0
E7	10	4	0	0	50	0
E8	0	0	0	0	3	0
E9	0	4	0	0	0	0
E10	0	0	0	0	1	0
E11	2	2	0	0	0	0
E12	10	0	0	0	100	0
E13	15	5	0	0	150	0
E14	3	0	0	0	4	0
E15	15	0	0	0	10	0
E16	0	0	0	1	0	0
E17	5	5	0	5	5	0
E18	20	0	0	10	10	0
E19	5	0	0	0	2	0
E20	2	0	0	0	5	0

Appendix 3e: Previous number of times that participants had used simulator training prior to enrolment where, N = Novice, I = Intermediate, A = Advanced, E = Expert

PARTICIPANT	FOUNDATION SKILLS	BASIC SKILLS	INT SKILLS	ADVANCED SKILLS	POLYPECTOMY
N1	0	0	0	0	0
N2	0	0	0	0	0
N3	0	0	0	0	0
N4	0	0	0	0	0
N5	0	0	0	0	0
N6	0	0	0	0	0
N7	0	0	0	0	0
N8	0	0	0	0	0
N9	0	0	0	0	0
N10	0	0	0	0	0
N11	0	0	0	0	0
N12	0	0	0	0	0
N13	0	0	0	0	0
N14	0	0	0	0	0
N15	0	0	0	0	0
N16	0	0	0	0	0
N17	0	0	0	0	0
N18	0	0	0	0	0
N19	0	0	0	0	0
N20	0	0	0	0	0
I1	0	1	0	0	0
I2	0	0	0	0	0
I3	0	0	0	0	0
I4	0	0	0	0	0
I5	0	0	0	0	0
I6	0	0	0	0	0
I7	0	1	0	0	0
I8	1	0	0	0	0
I9	1	1	0	0	0
I10	0	1	0	1	0
I11	1	1	0	0	0

I12	1	1	0	0	0
I13	0	1	0	0	0
I14	1	0	0	0	0
I15	1	0	0	0	0
I16	1	1	0	0	0
I17	1	1	0	0	0
I18	0	1	0	0	0
I19	1	1	0	0	0
I20	1	1	0	0	0
A1	0	1	0	1	0
A2	0	1	0	0	1
A3	0	1	0	0	0
A4	0	1	0	0	0
A5	1	1	1	1	1
A6	0	0	0	1	0
A7	1	0	1	0	0
A8	0	1	0	0	0
A9	1	1	0	0	0
A10	0	1	0	1	0
A11	0	0	0	0	0
A12	0	0	0	0	0
A13	1	0	1	1	1
A14	1	1	0	1	0
A15	1	1	0	1	1
A16	1	1	0	1	1
A17	1	1	1	0	1
A18	0	0	0	0	0
A19	1	1	0	0	0
A20	1	1	0	1	0
E1	1	1	1	1	1
E2	1	1	1	1	1
E3	1	1	1	1	1

E4	1	1	1	1	0
E5	1	1	0	1	1
E6	0	0	0	1	0
E7	1	1	1	1	1
E8	0	1	1	1	0
E9	0	0	1	0	0
E10	0	0	1	1	1
E11	0	1	0	0	1
E12	0	1	0	1	1
E13	1	1	1	1	1
E14	0	1	0	1	0
E15	0	1	1	1	1
E16	0	0	0	0	0
E17	1	1	1	1	1
E18	1	1	1	1	1
E19	1	1	0	0	0
E20	0	1	1	1	1

Appendix 3f: Previous number of courses attended by participants prior to construct study where, N = Novice, I = Intermediate, A = Advanced, E = Expert, 0 = course not completed, 1 = course completed

PARTICIPANT	COLONOSCOPY	FLEXI SIG	SESSILE	PEDUNCULATED
N1	0	0	0	0
N2	0	0	0	0
N3	0	0	0	0
N4	0	0	0	0
N5	0	0	0	0
N6	0	0	0	0
N7	0	0	0	0
N8	0	0	0	0
N9	0	0	0	0
N10	0	0	0	0
N11	0	0	0	0
N12	0	0	0	0
N13	0	0	0	0
N14	0	2	0	0
N15	0	0	0	0
N16	0	0	0	0
N17	0	0	0	0
N18	0	0	0	0
N19	0	0	0	0
N20	0	0	0	0
I1	20	30	3	5
I2	20	20	0	1
I3	225	400	30	50
I4	50	5	0	0
I5	50	50	5	5
I6	35	60	10	10
I7	55	65	5	5
I8	100	100	20	20
I9	60	80	5	3
I10	140	150	15	15
I11	20	50	0	1

I12	51	150	50	50
I13	60	80	0	0
I14	50	50	10	10
I15	60	80	2	5
I16	85	100	5	5
I17	150	200	10	10
I18	50	50	20	20
I19	50	100	3	4
I20	120	200	15	15
A1	500	200-500	41-60	>100
A2	1300	500	100	41-60
A3	>500	500	500	>500
A4	>500	500	500	>500
A5	>500	500	500	>500
A6	800	500	61-80	61-80
A7	500	500	500	>500
A8	300	300	81-100	81-100
A9	200-500	200-500	21-40	21-40
A10	500	200-500	21-40	41-60
A11	50	151-200	21-40	21-40
A12	500	500	100	>100
A13	500	500	81-100	41-60
A14	500	500	100	100
A15	2000	1500	10	40
A16	1100	550	50	70
A17	>500	500	100	81-100
A18	>500	500	100	>100
A19	1400	500	100	41-60
A20	1000	600	500	200
E1	>500	>500	>100	>100
E2	>500	>500	>100	>100
E3	5000	2000	>100	>100

E4	4000	2000	500	200
E5	>500	>500	>100	>100
E6	>500	>500	>100	>100
E7	4500	2500	400	250
E8	>500	>500	>100	>100
E9	>500	>500	>100	>100
E10	>500	>500	>100	81-100
E11	>500	>500	81-100	81-100
E12	>500	>500	>100	>100
E13	>500	>500	>100	>100
E14	2500	>500	>100	>100
E15	>500	>500	>100	>100
E16	>500	>500	>100	>100
E17	>500	>500	>100	>100
E18	>500	>500	>100	>100
E19	>500	>500	>100	>100
E20	>500	>500	>100	>100

Appendix 3g: Previous number of endoscopic procedures performed by participants where,

N = Novice, I = Intermediate, A = Advanced, E = Expert,

PARTICIPANT	DOPyS parameters measured via shortened DOPyS Assessment							
	OPTIMAL POSITION	OPTIMISING VIEW	POSITIONING SNARE	POSITIONING AT STALK	TRAPPING	DIATHERMY	RETRIEVAL	OVERALL
I1	2	2	2	2	2	2	2	2
I2	1	1	1	1	1	1	1	1
I3	1	1	1	1	1	1	1	1
I6	2	2	2	2	2	2	2	2
I7	2	2	2	2	2	2	2	2
I8	2	2	3	2	2	2	2	2
I9	2	2	2	2	2	2	2	2
I10	2	2	2	2	2	3	2	2
I13	1	1	1	1	1	1	1	1
I14	2	2	2	2	2	1	2	2
I15	1	1	1	1	1	1	1	1
I16	1	1	1	1	1	1	1	1
I17	2	2	2	2	2	2	2	2
I18	1	1	1	1	1	1	0	1
I19	1	1	1	1	1	1	1	1
I20	1	1	1	1	1	1	1	1

Appendix 3h: “Real-life” DOPyS scores submitted by the intermediate group (I)

PARTICIPANT	DOPyS parameters measured via shortened DOPyS Assessment							
	OPTIMAL POSITION	OPTIMISING VIEW	POSITIONING SNARE	POSITIONING AT STALK	TRAPPING	DIATHERMY	RETRIEVAL	OVERALL
N1	0	0	0	0	0	0	0	0
N2	1	1	1	1	1	1	1	1
N3	0	0	0	0	0	0	0	0
N4	1	1	1	1	1	1	1	1
N5	1	1	1	1	1	1	1	1
N6	1	1	1	1	1	1	1	1
N7	1	1	1	1	1	1	1	1
N8	0	0	0	0	0	0	0	0
N9	1	1	1	1	1	1	1	1
N10	0	0	0	0	0	0	0	0
N11	1	1	1	1	1	1	1	1
N12	0	0	0	0	0	0	0	0
N13	0	0	0	0	0	0	0	0
N14	0	0	0	0	0	0	0	0
N15	0	0	0	0	0	0	0	0
N16	1	1	1	1	1	1	1	1
N17	1	1	1	1	1	1	1	1
N18	0	0	0	0	0	0	0	0
N19	1	1	1	1	1	1	1	1

N20	0	0	0	0	0	0	0	0
I1	2	2	2	2	2	2	2	2
I2	1	1	1	1	1	1	1	1
I3	3	3	3	3	3	3	3	3
I4	0	0	0	0	0	0	0	0
I5	1	1	1	1	1	1	1	1
I6	2	2	2	2	2	2	2	2
I7	2	2	2	2	2	2	2	2
I8	3	3	3	3	3	3	3	3
I9	1	1	1	1	1	1	1	1
I10	2	2	2	2	2	2	2	2
I11	2	2	2	2	2	2	2	2
I12	3	3	3	3	3	3	3	3
I13	2	2	2	2	2	2	2	2
I14	2	2	2	2	2	2	2	2
I15	1	1	1	1	1	1	1	1
I16	1	1	1	1	1	1	1	1
I17	2	2	2	2	2	2	2	2
I18	2	2	2	2	2	2	2	2
I19	1	1	1	1	1	1	1	1
I20	1	1	1	1	1	1	1	1
A1	3	3	3	3	3	3	3	3
A2	3	3	3	3	3	3	3	3

A3	3	3	3	3	3	3	3	3
A4	3	3	3	3	3	3	3	3
A5	3	3	3	3	3	3	3	3
A6	3	3	3	3	3	3	3	3
A7	4	4	3	3	3	4	4	3
A8	3	3	3	3	3	3	3	3
A9	3	3	3	3	3	3	3	3
A10	3	3	3	3	3	3	3	3
A11	4	4	4	4	4	4	4	4
A12	3	3	3	3	2	3	2	3
A13	4	4	4	4	4	4	4	4
A14	3	3	3	3	3	3	3	3
A15	2	3	3	3	2	3	3	3
A16	3	3	3	3	3	3	3	3
17	3	3	3	3	3	3	3	3
A18	3	3	3	3	3	3	3	3
A19	3	3	3	3	3	3	3	3
A20	3	3	3	3	3	3	3	3
E1	4	4	4	4	4	4	4	4
E2	4	4	4	4	4	4	4	4
E3	4	4	3	3	3	3	4	4
E4	3	3	3	3	3	3	3	3
E5	3	3	3	3	3	3	3	3

E6	4	4	4	4	4	4	4	4
E7	4	4	4	4	4	4	4	4
E8	4	4	4	4	4	4	4	4
E9	3	3	3	3	3	3	3	3
E10	3	3	3	3	3	3	3	3
E11	3	3	3	3	3	3	3	3
E12	4	4	4	4	4	4	4	4
E13	4	4	4	4	4	4	4	4
E14	3	3	3	3	3	3	3	3
E15	4	4	4	4	4	4	4	4
E16	4	4	4	4	4	4	4	4
E17	4	4	4	4	4	4	4	4
E18	4	4	4	4	4	4	4	4
E19	3	3	4	4	4	3	4	4
E20	3	3	3	3	3	3	3	3

Appendix 3i: Self Assessment DOPyS scores submitted by the construct validation cohort. N = Novice, I = Intermediate, A = Advanced, E = Expert

ID	*DOPyS parameters measured via shortened DOPyS Assessment*																
	POA (1)	**POA (2)**	**VIA (1)**	**VIA (2)**	**SNA (1)**	**SNA (2)**	**STA (1)**	**STA (2)**	**TRA (1)**	**TRA (2)**	**DIA (1)**	**DIA (2)**	**REA (1)**	**REA (2)**	**OVA (1)**	**OVA (2)**	**TIME**
N1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	635
N2	1	1	1	2	2	1	2	1	2	1	2	2	2	2	2	2	212
N3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	338
N4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	527
N5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	376
N6	1	1	1	1	2	1	2	1	2	1	2	1	2	1	2	1	382
N7	2	2	2	2	3	2	3	2	2	2	2	2	3	2	2	2	217
N8	3	3	2	3	3	3	2	3	2	3	2	3	3	3	2	3	230

N9	2	1	2	1	2	2	2	1	1	1	1	2	1	1	2	2	441
N10	1	2	1	2	1	2	1	2	1	2	1	2	1	1	1	2	417
N11	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1	1080
N12	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	684
N13	1	1	1	1	0	1	0	1	0	1	0	1	0	1	0	1	620
N14	2	1	1	1	1	1	2	1	2	1	1	1	0	1	1	1	617
N15	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	255
N16	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	883
N17	1	1	1	1	1	1	0	1	0	1	0	1	0	1	1	1	1263
N18	3	2	2	1	2	2	1	1	1	2	1	2	1	1	1	2	1343
N19	2	2	1	1	1	1	1	2	1	2	1	1	0	0	1	1	512

N20	1	1	2	2	2	2	1	1	1	2	2	1	1	2	1	2	317
I1	2	2	2	2	2	2	2	2	3	2	3	2	3	3	2	2	323
I2	3	1	3	1	1	1	1	1	1	1	1	1	1	1	1	1	712
I3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	249
I4	2	3	2	3	3	3	3	3	3	3	3	3	2	3	3	3	324
I5	1	3	2	3	3	3	2	3	3	3	2	3	3	3	2	3	191
I6	2	3	2	3	1	3	2	3	2	3	2	3	2	3	2	3	451
I7	1	3	1	3	1	3	1	3	1	3	1	3	2	3	1	3	333
I8	3	2	2	2	2	2	2	2	2	2	3	2	3	2	2	2	415
I9	2	2	2	2	2	2	2	2	2	2	2	2	3	2	2	2	241
I10	3	1	3	2	1	2	1	1	1	2	1	2	2	2	1	2	306

I11	1	3	1	3	2	3	2	3	2	3	2	3	3	3	2	3	248
I12	3	2	2	2	2	2	2	2	3	2	2	2	3	2	2	2	489
I13	2	2	1	2	2	2	2	2	2	2	2	3	3	3	2	2	252
I14	3	3	2	3	2	3	2	3	2	3	2	3	3	3	2	3	307
I15	3	2	3	2	2	3	2	3	3	2	3	2	3	2	3	2	315
I16	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	206
I17	3	2	3	3	3	3	4	2	4	3	3	3	3	3	3	3	197
I18	4	4	4	4	4	3	4	3	4	3	4	4	4	4	4	3	125
I19	1	3	1	3	1	3	1	2	2	2	1	3	1	3	1	3	203
I20	2	2	2	2	2	2	3	2	3	2	3	2	3	2	3	2	554
A1	3	3	4	3	4	3	4	4	4	4	4	3	3	3	4	3	242

A2	3	2	3	2	3	3	3	3	3	3	2	2	3	3	3	3	301
A3	2	2	2	2	1	1	1	1	2	1	2	1	1	2	2	2	598
A4	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	423
A5	3	2	3	2	3	3	3	2	3	3	3	2	3	3	3	3	255
A6	3	4	3	4	3	3	1	3	3	3	2	4	2	4	3	3	274
A7	2	2	2	2	3	3	3	2	3	2	3	2	2	2	3	2	298
A8	3	3	3	3	3	3	2	3	4	3	4	3	4	4	3	3	198
A9	4	3	4	3	3	3	4	3	3	3	3	3	4	4	4	3	272
A10	3	4	3	4	2	4	2	4	2	4	2	4	2	4	2	4	168
A11	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	166
A12	3	4	2	4	3	3	3	3	3	3	3	3	4	4	3	3	193

A13	1	4	2	4	1	3	1	3	2	3	2	3	1	4	2	3	557
A14	2	3	1	3	2	3	1	3	1	3	2	3	2	2	2	3	305
A15	4	3	4	4	3	3	3	4	3	3	4	4	4	3	4	4	262
A16	3	4	3	4	3	4	3	4	4	4	4	4	4	4	3	4	244
A17	3	4	3	4	3	4	3	4	4	4	4	4	4	4	4	4	264
A18	3	3	3	3	3	4	3	3	3	3	3	4	4	4	3	3	233
A19	3	4	3	4	3	4	3	4	3	4	3	4	3	4	3	4	404
A20	3	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	435
E1	3	2	3	2	3	3	3	3	3	3	4	2	3	3	3	3	443
E2	3	2	2	2	3	2	2	2	2	2	3	2	1	1	3	3	184
E3	3	2	3	2	3	2	3	2	3	2	3	2	2	2	3	2	578

E4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	377
E5	3	4	3	4	3	3	2	3	3	3	2	3	3	4	3	4	495
E6	4	3	4	3	3	3	4	3	4	3	4	3	3	3	4	3	460
E7	3	3	4	2	2	3	3	2	4	3	2	3	4	3	3	3	330
E8	4	3	4	3	3	3	2	3	3	3	3	3	3	4	3	3	315
E9	4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	355
E10	4	4	4	4	4	4	4	4	4	4	4	4	3	4	4	4	179
E11	2	3	1	3	2	3	3	3	3	3	3	3	3	4	3	3	332
E12	4	4	3	4	4	4	4	4	4	4	4	4	4	4	4	4	145
E13	3	3	3	3	3	3	3	3	3	3	3	3	3	4	3	3	244
E14	4	2	3	2	2	2	2	2	2	2	2	2	2	2	2	2	254

E15	4	4	2	4	4	4	4	4	4	4	4	4	4	4	4	4	161
E16	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	257
E17	3	3	2	3	3	3	3	3	3	3	2	3	3	3	3	3	178
E18	3	4	2	4	3	4	3	4	3	4	3	4	2	4	3	4	389
E19	2	3	1	3	2	3	3	3	3	3	3	3	2	3	2	3	454
E20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	713

Appendix 3j: DOPyS results for polyp A task for assessors (1) and (2) for construct validation study. PO = Attempts to achieve optimal polyp position, VI = Optimises view by aspiration/ insufflation/ wash, SN = Directs snare accurately over polyp head, ST = Advances snare sheath towards stalk as closed, TR = Appropriate amount of tissue trapped in snare, DI = Applies appropriate degree of diathermy, RE = Retrieves or attempts to retrieve polyp, OV = Overall competency at polypectomy. Time taken in Seconds. N = Novice, I = Intermediate, A = Advanced, E = Expert.

ID	*DOPyS parameters measured via shortened DOPyS Assessment*																
	VIB (2)	**POB (1)**	**POB (2)**	**VIB (1)**	**SNB (1)**	**SNB (2)**	**STB (1)**	**STB (2)**	**TRB (1)**	**TRB (2)**	**DIB (1)**	**DIB (2)**	**REB (1)**	**REB (2)**	**OVB (1)**	**OVB (2)**	**TIME**
N1	0	1	0	1	1	0	0	0	1	0	1	0	0	0	1	0	720
N2	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	553
N3	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	567
N4	1	2	1	2	1	1	0	1	0	1	0	1	0	1	0	1	486
N5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	467
N6	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	1	706
N7	1	2	1	1	1	1	0	1	0	1	0	1	0	1	0	1	867
N8	3	3	3	2	3	3	2	3	2	3	2	3	1	3	2	3	1204
N9	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1603
N10	1	1	1	1	1	1	1	2	1	2	1	2	1	1	1	1	1104
N11	0	1	0	1	1	0	1	0	1	0	1	0	1	0	1	0	521
N12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done

N13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done
N14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done
N15	1	0	1	0	0	1	0	1	0	1	0	1	0	1	0	1	664
N16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done
N17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done
N18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	612
N19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	not done
N20	1	2	1	2	1	1	1	1	1	1	1	1	1	1	1	1	702
I1	1	1	1	2	1	1	0	1	0	1	0	1	0	0	0	1	698
I2	1	1	1	2	1	1	0	1	0	1	0	1	0	1	0	1	830
I3	2	2	3	2	0	2	0	2	0	3	0	2	0	2	0	2	398
I4	3	2	3	2	2	3	2	3	2	3	2	3	2	3	2	3	352
I5	2	2	3	2	1	2	2	2	2	2	2	3	2	3	2	3	599
I6	3	3	3	2	2	3	3	3	3	3	3	3	3	3	3	3	481

I7	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	361
I8	1	0	1	0	0	2	0	2	0	0	0	0	0	0	0	0	770
I9	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2	2	1140
I10	2	1	2	1	2	1	2	2	2	2	2	2	2	2	2	2	1052
I11	3	2	3	2	2	3	2	3	2	3	2	3	2	3	2	3	473
I12	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3	3	516
I13	1	2	1	1	1	2	1	2	1	2	1	2	1	1	1	2	667
I14	2	3	2	2	1	2	1	2	1	2	1	2	1	2	1	2	560
I15	1	2	1	1	2	1	3	1	3	1	3	1	3	1	3	1	540
I16	3	2	3	1	2	3	3	3	3	3	3	3	3	3	3	3	684
I17	2	1	2	2	2	3	2	2	2	2	2	3	2	2	2	2	503
I18	3	3	3	3	3	3	4	3	4	3	4	2	4	3	4	3	340
I19	3	1	3	1	1	3	1	2	1	2	1	3	1	3	1	3	952
I20	2	1	2	2	2	2	3	2	3	2	3	2	3	2	3	2	371
A1	3	4	3	4	3	2	3	3	4	2	4	3	4	3	4	3	271
A2	3	3	3	3	4	3	4	3	4	3	3	3	3	3	3	3	280

A3	2	2	2	2	2	2	2	1	1	1	1	1	1	2	2	2	482
A4	2	2	2	2	1	2	1	2	1	2	2	2	2	2	2	2	598
A5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	424
A6	3	3	3	2	1	3	3	4	3	4	3	3	2	4	3	4	308
A7	1	2	1	2	1	1	1	1	1	1	1	2	1	0	1	1	421
A8	2	3	2	2	3	2	4	2	4	2	4	2	3	2	3	2	447
A9	3	3	3	3	3	3	4	3	4	3	4	3	3	3	3	3	479
A10	2	2	2	2	1	2	2	2	2	2	2	2	0	2	2	2	443
A11	3	4	3	4	3	3	4	4	4	4	4	4	4	3	4	4	218
A12	3	2	3	3	3	3	1	3	1	3	1	3	2	2	2	3	241
A13	3	2	3	3	3	3	3	3	1	3	1	3	2	3	2	3	212
A14	3	2	3	2	3	3	2	3	2	3	2	3	3	3	2	3	245
A15	3	2	3	1	2	3	2	3	2	3	1	3	0	3	2	3	384
A16	3	3	3	3	2	3	3	3	3	3	3	3	2	2	3	3	804
A17	3	4	3	4	3	3	3	3	4	3	4	3	4	3	4	3	744
A18	3	4	3	3	3	3	4	3	4	3	4	3	3	3	4	3	713

A19	2	3	2	2	3	3	2	3	2	3	3	3	1	2	2	3	1194
A20	2	2	2	1	1	2	1	2	1	2	1	2	1	2	1	2	803
E1	3	3	2	3	2	2	3	3	2	2	2	2	3	2	3	3	482
E2	2	3	2	2	3	2	2	2	3	2	2	2	2	2	3	2	280
E3	2	3	2	3	3	2	2	2	3	2	2	2	2	2	3	2	391
E4	4	2	4	2	2	4	2	4	2	4	2	4	2	4	2	4	680
E5	3	3	3	2	2	3	3	3	3	3	3	3	3	2	3	3	500
E6	3	4	3	3	2	3	2	3	2	3	2	3	2	3	2	3	413
E7	3	3	3	4	4	3	4	3	2	3	4	3	2	3	4	3	604
E8	4	4	4	3	3	4	2	4	3	4	3	4	3	4	3	4	275
E9	1	3	2	2	2	1	2	2	2	1	2	2	2	2	2	2	920
E10	4	3	4	3	3	4	4	4	4	4	4	4	3	4	4	4	354
E11	3	3	3	3	2	3	2	3	3	3	3	3	3	3	3	3	501
E12	4	3	4	3	4	4	4	4	4	4	4	4	4	4	4	4	405
E13	3	2	3	3	2	3	2	3	2	3	3	3	2	3	3	3	368
E14	3	4	3	3	4	3	3	3	4	3	4	3	4	3	4	3	270

E15	4	3	4	2	4	4	4	4	4	4	4	4	4	4	4	4	156
E16	4	3	4	3	3	4	4	4	4	4	4	4	3	4	3	4	387
E17	3	2	3	2	2	3	3	3	2	3	2	3	2	3	2	3	183
E18	4	2	4	2	3	4	2	4	2	4	2	4	3	4	2	4	532
E19	3	2	3	2	2	4	3	4	3	3	3	4	2	3	2	3	396
E20	1	2	1	1	1	1	2	1	2	1	2	1	1	1	2	1	881

Appendix 3k: DOPyS results for polyp B task for assessors (1) and (2) for construct validation study. PO = Attempts to achieve optimal polyp position, VI = Optimises view by aspiration/ insufflation/ wash, SN = Directs snare accurately over polyp head, ST = Advances snare sheath towards stalk as closed, TR = Appropriate amount of tissue trapped in snare, DI = Applies appropriate degree of diathermy, RE = Retrieves or attempts to retrieve polyp, OV = Overall competency at polypectomy. Time taken in Seconds. N = Novice, I = Intermediate, A = Advanced, E = Expert.

Appendix 4:

Data from Skills Transfer Study

School of Postgraduate Medical and Dental Education
Cardiff University
Neuadd Meirionnydd
Heath Park
Cardiff CF14 4YS

Tel *Ffôn* +44 (0)29 2068 7441
Fax *Ffacs* + 44 (0)29 2068 7455
E-mail *E-bost* pugsleyla@cf.ac.uk

Ysgol Addysg Feddygol a Deintyddol ôl Raddedig
Prifysgol Caerdydd
Neuadd Meirionnydd
Mynydd Bychan
Caerdydd CF14 4YS

23/07/2014

Dear James

Project title: A Prospective Randomised Trial Comparing Ex Vivo Simulation Training with Virtual Reality Simulation training in Colonoscopic Polypectomy.

This is to confirm that the proposal for the above research study, together with your completed Research Ethics declaration form, has been considered by the Research Ethics Committee for the School of Postgraduate Medical and Dental Education. The Committee feel that this study will take the form of an educational evaluation and that full awareness of research ethics guidelines has been shown.

Decision: The Committee confirms that Ethical Approval is granted for this study

Dr Lesley Pugsley

Senior Lecturer in Medical Education

Chair of PGMDE Research Ethics Committee

Appendix 4a: Approval Letter from the Research Ethics Committee for the School of Postgraduate Medical and Dental Education for Skills transfer RCT

Research and Commercial Division
Director Geraint W Jones
Adran Ymchwil a Masnach
Cyfarwyddwr Geraint W Jones

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Cardiff University
7th Floor
30 - 36 Newport Road
Cardiff CF24 0DE
Wales UK
Tel *Ffôn* +44(0)29 2087 5834
Fax *Ffacs* +44(0)29 2087 4189
Prifysgol Caerdydd
Llawr 7
30 - 36 Heol Casnewydd
Caerdydd CF24 0DE
Cymru Y Deyrnas Gyfunol

18th April 2012

Mr Jared Torkington
Associate Dean of Simulation/ Consultant Colorectal Surgeon
The Wales Deanery/ Department of Colorectal Surgery
University Hospital of Wales
Heath Park
Cardiff
CF14 4XN

Dear Dr Torkington,

A prospective, randomised trial comparing simulator training to clinical training in colonoscopic polypectomy

I understand that you are acting as Academic Supervisor for the above MD project to be conducted by Mr James Ansell.

I confirm that Cardiff University agrees to act as Sponsor for the above project, as required by the Research Governance Framework for Health and Social Care. I can also confirm that Scientific Review has been obtained from Dr Pete Wall (ISCA Healthcare Research) and Dr Neil Warren (Senior Lecturer in Simulation, Welsh Institute for Minimal Access Therapy).

The necessary insurance provisions will be in place prior to the project commencement. Cardiff University is insured with Zurich Municipal. Copies of the insurance certificate are attached to this letter.

Pleae note that you will need to apply (via IRAS) to NISCHR Permissions Coordinating Unit (PCU) if you decide to recruit Wales-based NHS staff through their place of work. For NHS staff recruited from Trusts outside of Wales, permission will need to be sought through the relevant R&D office.

May I take this opportunity to remind you that, as Chief Investigator, you are required to:

- ensure you are familiar with your responsibilities under the Research Governance Framework for Health and Social Care;
- undertake the study in accordance with Cardiff University's Research Governance Framework and the principles of Good Clinical Practice;
- ensure the Research complies with the Data Protection Act 1998;
- inform the Research and Commercial Division (RACD) of any amendments to the protocol or study design, including changes to start /end dates;
- co-operate with any audit inspection of the project files or any requests from RACD for further information.

You should quote the following unique reference number in any correspondence relating to sponsorship for the above project:

SPON 1100-12

This reference number should be quoted on all documentation associated with this project.

Yours sincerely

H. Falcaro

pp. **Dr K J Pittard Davies**
Head of Research Policy & Management
Direct line: +44 (0) 29208 79274
Email: DaviesKP2@cf.ac.uk

cc Mr James Ansell

Appendix 4b: Approval Sponsorship letter from the Research and Commercial Division, Cardiff University to conduct Skills transfer RCT

Postgraduate Dean and Head of School
Professor Derek Gallen FRCGP FHEA MMED FAcadMEd FRCP
Deputy Postgraduate Dean
Professor Peter Donnelly FRCPsych BA (Open) PGCME ILTM
Pennaeth yr Ysgol a Deon
Yr Athro Derek Gallen FRCGP FHEA MMED FAcadMEd FRCP
Dirprwy Deon
Yr Athro Peter Donnelly FRCPsych BA (Open) PGCME ILTM

Wales Deanery Deoniaeth Cymru

CARDIFF UNIVERSITY PRIFYSGOL CAERDYDD

wimat

WIMAT
Cardiff Medicentre
Heath Park
Cardiff
CF14 4UJ

RE: A Prospective Randomised Trial comparing Ex-vivo Animal simulation training with Virtual Reality simulation training in colonic polypectomy

Dear Consultant Trainer,

This trainee is currently participating in a clinical trial to evaluate the usefulness of the WIMAT colonoscopy suitcase. This is a simulator designed to teach the skill of colonic polypectomy. The trainee has taken part in standardised polypectomy skills training on either the WIMAT colonoscopy suitcase or a control. We are measuring the impact this training has on their ability to perform "real-life" polypectomy. The trial has been approved by the Research and Commercial Division at Cardiff University and the Post Graduate Deanery Wales Ethics committee.

We would be extremely grateful if the trainee is given the opportunity to perform colonic polypectomy and DOPyS assessment if you deem this appropriate for their current level of training. Further information regarding the trial can be found at:

http://controlled-trials.com/ISRCTN41736707/

If you have any questions regarding the research please contact us at the addresses provided below. Many thanks for your support.

Kind regards,

Mr Jared Torkington
Consultant Colorectal Surgeon,
University Hospital of Wales,
Heath Park,
Cardiff.
Jared.torkington@wales.nhs.uk
02920 682131

Mr James Ansell
RCSEng Clinical Research Fellow,
Welsh Institute for Minimal Access Therapy,
Cardiff Medicentre,
Cardiff.
ansellj@cf.ac.uk
02920 682138

Appendix 4c: Letter to consultants trainers to notify that trainees are enrolled in skills transfer

Randomised Controlled Trial

Transfer of technical skills study						
Age	_____					
Sex		Male ☐			Female ☐	
Grade	CT1-2 ☐	ST3 ☐	ST4 ☐	ST5 ☐	ST6 ☐	ST7 ☐
Hospital						
Specialty						
How many flexible sigmoidoscopies have you performed to date?	0-5 ☐	5-10 ☐	11-15 ☐	16-20 ☐	21-25 ☐	>25 ☐
How many colonoscopies have you performed to date?	0-5 ☐	5-10 ☐	11-15 ☐	16-20 ☐	21-25 ☐	>25 ☐
How many OGDs have you performed to date?	0-5 ☐	5-10 ☐	11-15 ☐	16-20 ☐	21-25 ☐	>25 ☐
How many colonic polypectomies have you performed to date?	0 ☐	1-5 ☐	6-10 ☐	11-15 ☐	16-20 ☐	21-25 ☐
What colonoscopy Simulators have you use before?	Bench model ☐					
	Virtual reality ☐					
	Cadaveric ☐					
	Ex-vivo animal ☐					
	Live animal ☐					
What specialty will you be in from August 2012 to August 2013?						
Further comments:						

Appendix 4d: Proforma used to establish inclusion and exclusion criteria for skills transfer study

CONSENT FORM

THE VALIDATION OF THE COLONIC POLYPECTOMY SIMULATION

By signing this form you are agreeing to participate in a randomised controlled trial which is looking at the transfer of polypectomy skills from simulation to "real life". You will perform several assessed polypectomies on your assigned simulator. Your performance may be videoed and assessed. Assessors will be blinded at all times to your identity. You will need to submit ≥4 real life DOPyS assessment forms following.

For the participant (please initial each box):

1. ***I understand the nature of this trial and the details of my involvement have been fully explained*** ☐
2. ***I understand that I can withdraw from this trial at any stage*** ☐
3. ***I understand that the data collected will remain anonymous and stored on password secure university computers and permanently erased at the end of the trial*** ☐

SIGNED: ……

PRINT: ……..

DATE: ……..

For the chief investigator (please initial each box):

1. ***I certify that I have fully explained the nature of the trial and the details of the participants involvement*** ☐
2. ***I have explained that they are able to withdraw from the trial at any stage*** ☐
3. ***I have explained that the data collected will remain anonymous and stored on password secure university computers and permanently erased at the end of the trial*** ☐

SIGNED: ……

PRINT: ……..

DATE: ……..

Appendix 4e: Consort form using for skills transfer study

Name: _____ Hospital: _____ Level: _____	
How many colonoscopy training lists have you participated in since enrolling in the trial?	
How many OGD training lists have you participated in since enrolling in the trial?	
How many flexible sigmoidoscopies have you performed since enrolling in the trial?	
How many colonoscopies have you performed to since enrolling in the trial?	
How many OGDs have you performed to since enrolling in the trial?	
How many colonic polypectomies have you performed since enrolling in the trial?	
Have you attended any endoscopy courses since enrolling in the trial?	
Have you used any colonoscopy simulator (other than the WIMAT colonoscopy) suitcase since enrolling in the trial?	
Further comments:	

Appendix 4f: Proforma for collection of post intervention clinical experience

ID	Colon lists	OGD lists	No. Colons	No. Flexi	No. OGD	No. Polyps	Simulator training
2	0	0	0	0	0	0	0
4	34	4	52	154	15	4	0
16	7	2	26	21	97	7	0
7	5	6	5	41	10	4	0
8	8	4	25	25	30	3	0
9	0	0	5	5	5	1	0
10	10	10	30	10	40	4	0
20	0	0	0	0	0	0	0
1	2	8	6	3	30	0	0
5	2	0	6	6	2	0	0
13	5	0	5	10	0	1	0
14	5	0	5	20	2	0	0
15	3	2	12	38	6	4	0
18	5	0	9	7	1	0	0
19	0	10	0	0	35	0	0

Appendix 4g: Previous numbers of endoscopic procedures performed by cohort in the Skills transfer RCT (Grey-shaded Area representing Group B, unshaded Group A)

ID	FOUNDATION	BASIC COLON	BASIC THERAPEUTIC UGI	BASIC UGI	INT SKILLS	ADVANCED	POLYPECTOMY
2	**1**	0	0	0	0	0	0
4	0	1	1	0	0	0	0
16	1	1	1	1	0	0	0
7	1	1	1	1	0	0	0
8	0	1	0	1	0	0	0
9	1	0	0	1	0	0	0
10	1	1	0	1	0	0	0
20	0	0	0	0	0	0	0
1	1	0	0	1	0	0	0
5	1	0	0	0	0	0	0

11	0	0	1	1	0	0	0
13	1	0	0	0	0	0	0
14	1	0	0	0	0	0	0
15	1	1	0	0	0	0	0
18	0	1	0	0	0	0	0
19	0	1	0	1	0	0	0

Appendix 4h: Previous JAG accredited courses completed by cohort in the Skills transfer RCT (0 = not attended, 1= attended). (Grey-shaded Area representing Group B, unshaded Group A)

ID	Speciality	Bench	VR	Cadaveric	Ex-vivo	Live Animal	other
2	Breast and General	1	1	0	1	0	0
4	Colorectal	1	1	0	0	0	0
16	Gastroenterology	1	0	0	0	0	0
7	UGI	1	0	0	0	0	0
8	Colorectal Surgery	1	0	0	0	0	0
9	Colorectal Surgery	1	0	0	1	0	0
10	Gastroenterology	1	1	0	1	0	0
20	General/ Endocrine Surgery	1	0	0	0	0	0
1	Gastroenterology	0	0	0	0	0	0
5	Colorectal	0	0	0	0	0	0
11	Gastroenterology	1	0	0	0	0	0

13	Colorectal	0	0	0	0	0	0
14	Colorectal Surgery	0	0	0	0	0	0
15	Colorectal Surgery	0	1	0	0	0	0
18	General Surgery	1	0	0	0	0	0
19	UGI	1	3	0	1	0	0

Appendix 4i: Previous simulator experience by cohort in the Skills transfer RCT (0 = not attended, 1= attended). (Grey-shaded Area representing Group B, unshaded Group A)

ID	Flexi	Colon	OGD	Sessile	Pedun
2	5	0	5	0	0
4	79	58	25	0	0
16	200	170	100	20	10
7	37	21	550	0	1
8	45	75	200	6	7
9	10	0	15	0	0
10	50	110	100	0	10
20	55	40	20	13	8
1	39	107	229	22	3
5	6	2	2	0	0
11	100	100	1000	4	4
13	40	30	20	0	1
14	0	0	200	0	0
15	250	80	45	70	45
18	23	3	8	0	0
19	60	60	255	0	14

Appendix 4j: Previous numbers of endoscopic procedures performed cohort enrolled in the Skills transfer RCT.

(Grey-shaded Area representing Group B, unshaded Group A)

Participant	Training Session	**PO (1)**	**PO (2)**	**VI (1)**	**VI (2)**	**SN (1)**	**SN (2)**	**ST (1)**	**ST (2)**	**TR (1)**	**TR (2)**	**DI (1)**	**DI (2)**	**RE (1)**	**RE (2)**	**OV (1)**	**OV (2)**
2	1st	2	2	1	2	1	1	1	2	1	2	2	2	1	0	1	2
2	2nd	1	2	1	2	1	1	1	2	1	1	1	2	2	0	1	2
2	3rd	2	2	2	2	2	3	2	2	3	3	2	3	1	3	1	2
4	1st	2	3	2	3	2	3	3	3	3	3	2	3	1	3	2	3
4	2nd	2	4	2	4	2	3	3	4	3	3	2	3	2	3	2	3
4	3rd	3	3	2	4	3	4	3	4	3	4	3	4	2	3	3	4
16	1st	4	4	3	4	2	3	3	3	3	3	3	3	2	3	3	3
16	2nd	2	3	2	3	1	3	2	3	2	3	2	3	3	3	3	3
16	3rd	3	3	1	3	2	3	2	3	2	3	2	3	2	2	3	3
17	1st	3	3	2	3	3	3	3	3	3	4	3	4	2	3	2	3
17	2nd	2	3	1	3	1	3	2	4	3	3	1	3	1	3	1	3
17	3rd	4	4	2	4	2	3	2	3	3	3	2	3	2	4	2	3
7	1st	2	3	2	2	2	3	2	3	2	3	2	4	3	4	2	3
7	2nd	1	2	2	2	2	3	2	2	2	3	2	3	2	2	2	2
7	3rd	1	2	3	3	2	3	1	3	2	3	2	3	2	3	2	3

8	1st	2	2	2	2	0	2	0	0	0	0	0	0	0	0	0	0
8	2nd	2	2	2	3	2	3	3	3	3	3	2	3	2	3	3	3
8	3rd	2	3	2	2	2	3	2	3	2	3	2	3	2	3	3	3
9	1st	2	3	1	3	1	3	3	3	3	3	3	3	2	2	2	3
9	2nd	1	2	2	3	2	3	1	3	1	3	2	3	1	3	1	3
9	3rd	2	2	2	3	2	3	3	3	3	3	3	3	2	3	2	3
10	1st	4	4	3	4	4	4	3	3	3	3	3	3	3	3	3	3
10	2nd	3	4	3	4	4	4	3	3	3	3	3	3	4	4	3	4
10	3rd	2	2	2	3	2	1	1	1	1	1	1	2	1	0	2	2
20	1st	1	2	2	2	2	2	1	2	1	3	1	3	1	3	1	2
20	2nd	3	4	2	4	3	3	3	3	1	3	2	3	1	3	2	3
20	3rd	3	4	3	3	3	3	3	3	3	3	3	3	3	4	3	3
1	1st	3	2	3	2	3	3	2	3	3	3	3	3	2	3	3	3
1	2nd	2	2	2	2	3	3	2	2	3	3	3	3	3	3	3	2
1	3rd	3	3	3	3	3	3	3	2	3	2	3	3	3	3	3	3
5	1st	1	2	1	2	1	2	1	1	1	1	1	2	2	2	1	2
5	2nd	1	2	1	3	1	3	2	3	2	3	2	3	1	3	1	3

5	3rd	3	3	2	3	2	3	1	2	1	3	2	3	1	3	2	3
11	1st	4	3	4	3	4	3	4	3	4	3	4	3	2	0	4	3
11	2nd	1	2	2	3	2	3	2	3	2	3	3	3	3	3	2	3
11	3rd	3	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3
13	1st	2	3	2	3	2	2	3	2	3	2	2	3	3	3	2	2
13	2nd	3	2	3	3	2	3	1	2	2	3	2	3	1	0	2	2
13	3rd	2	3	1	3	2	3	1	3	1	3	1	3	2	3	1	3
14	1st	3	3	2	3	2	3	2	3	3	3	2	3	3	3	3	3
14	2nd	1	3	1	3	2	3	1	3	1	3	2	3	1	0	1	3
14	3rd	3	3	2	3	2	3	3	3	3	3	3	3	3	3	3	3
15	1st	2	3	2	3	1	3	3	3	3	3	2	3	3	4	2	3
15	2nd	2	3	2	3	1	3	2	3	3	3	3	3	2	3	2	3
15	3rd	1	2	2	2	1	3	1	3	1	3	1	3	3	3	1	3
18	1st	3	4	3	4	2	3	3	3	3	3	3	3	3	4	3	3
18	2nd	3	4	3	4	3	4	3	3	3	3	3	3	2	3	3	3
18	3rd	4	4	3	4	4	4	3	4	3	4	3	4	4	4	3	4
19	1st	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

19	2nd	1	2	2	3	2	3	2	3	3	3	3	3	2	3	2	3
19	3rd	1	2	1	1	1	1	0	0	0	0	0	0	0	0	0	1

Appendix 4k: DOPyS scores from Assessors (1) and (2) for 3 sessions of simulator training. PO = Attempts to achieve optimal polyp position, VI = Optimises view by aspiration/ insufflation/ wash, SN = Directs snare accurately over polyp head, ST = Advances snare sheath towards stalk as closed, TR = Appropriate amount of tissue trapped in snare, DI = Applies appropriate degree of diathermy, RE = Retrieves or attempts to retrieve polyp, OV = Overall competency at polypectomy. Time taken in Seconds. N = Novice, I = Intermediate, A = Advanced, E = Expert.

(Grey-shaded Area representing Group B, unshaded Group A)

Participant	DOPyS parameter assessed																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	3	1	3	2	2	1	3	2	3	0	2	2	0	3	0	3	3	3	0	0	3	3	0	2
16	4	4	3	4	4	4	3	3	3	4	3	4	3	3	4	3	4	3	4	0	0	3	4	0	4
7	2	2	2	2	3	2	3	2	2	2	0	3	3	3	2	3	2	3	3	3	0	3	3	0	2
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	3	3	3	3	3	3	3	3	3	4	4	3	3	4	3	3	3	4	3	3	3	3	3	3	3
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	4	4	3	3	3	3	4	3	3	3	0	3	3	0	3	3	3	3	3	0	0	2	3	3	3

15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 4l: DOPyS scores from the **first** "real-life" polyp performed following simulator polypectomy training. (Grey-shaded Area representing Group B, unshaded Group A). 1 = Attempts to achieve optimal polyp position, 2 = Optimises view by aspiration / insufflation / wash, 3 = Determines the full extent of the lesion, 4 = Uses appropriate polypectomy technique, 5 = Adjusts / stabilises scope position, 6 = Checks all polypectomy equipment available, 7 = Checks snare closure prior to introduction into the scope, 8 = Clear instructions to and utilisation of endoscopy staff, 9 = Checks diathermy settings are appropriate, 10 = Photo-documents pre and post polypectomy, 11 = Applies prophylactic haemostatic measures if appropriate, 12 = Selects appropriate sized snare, 13 = Directs snare accurately over polyp head, 14 = Correctly selects en-bloc/ piecemeal removal depending on size, 15 = Advances snare sheath towards stalk as snare closed, 16 = Places snare at appropriate position on, 17 = Mobilises polyp to ensure appropriate amount of tissue trapped, 18 = Applies appropriate degree of diathermy, 19 = Examines remnant stalk/ polyp base, 20 = Identifies and appropriately treats residual polyp, 21 = Identifies bleeding and performs adequate endoscopic haemostasis, 22 = Retrieves, or attempts retrieval of polyp, 23 = Checks for polyp retrieval, 24 = Places tattoo completely, where appropriate, 25 = Overall competency at Polypectomy

Participant	DOPyS parameter assessed																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	3	2	3	3	2	2	3	3	3	2	3	3	2	3	3	3	3	3	2	2	3	3	2	2
16	3	4	3	4	4	3	0	4	3	4	0	3	4	4	3	3	4	3	4	0	0	4	4	0	4
7	4	4	4	4	4	4	4	4	4	4	0	4	4	4	4	4	4	4	0	4	0	4	4	0	2
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	4	4	3	3	3	4	4	3	3	3	0	3	3	0	3	3	3	3	3	0	0	2	3	3	3

15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 4m: DOPyS scores from the **second** "real-life" polyp performed following simulator polypectomy training. (Grey-shaded Area representing Group B, unshaded Group A). 1 = Attempts to achieve optimal polyp position, 2 = Optimises view by aspiration / insufflation / wash, 3 = Determines the full extent of the lesion, 4 = Uses appropriate polypectomy technique, 5 = Adjusts / stabilises scope position, 6 = Checks all polypectomy equipment available, 7 = Checks snare closure prior to introduction into the scope, 8 = Clear instructions to and utilisation of endoscopy staff, 9 = Checks diathermy settings are appropriate, 10 = Photo-documents pre and post polypectomy, 11 = Applies prophylactic haemostatic measures if appropriate, 12 = Selects appropriate sized snare, 13 = Directs snare accurately over polyp head, 14 = Correctly selects en-bloc/ piecemeal removal depending on size, 15 = Advances snare sheath towards stalk as snare closed, 16 = Places snare at appropriate position on, 17 = Mobilises polyp to ensure appropriate amount of tissue trapped, 18 = Applies appropriate degree of diathermy, 19 = Examines remnant stalk/ polyp base, 20 = Identifies and appropriately treats residual polyp, 21 = Identifies bleeding and performs adequate endoscopic haemostasis, 22 = Retrieves, or attempts retrieval of polyp, 23 = Checks for polyp retrieval, 24 = Places tattoo completely, where appropriate, 25 = Overall competency at Polypectomy

Participant	DOPyS parameter assessed																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	3	3	3	3	3	3	3	3	3	2	3	3	2	3	3	3	3	3	2	3	3	3	2	2
16	3	3	3	4	3	3	0	3	3	4	0	3	3	3	3	3	4	3	3	0	0	3	3	0	3
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	4	3	3	3	3	4	4	3	3	4	0	3	3	0	3	2	2	3	3	0	0	3	3	3	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 4n: DOPyS scores from the **third** "real-life" polyp performed following simulator polypectomy training. (Grey-shaded Area representing Group B, unshaded Group A). 1 = Attempts to achieve optimal polyp position, 2 = Optimises view by aspiration / insufflation / wash, 3 = Determines the full extent of the lesion, 4 = Uses appropriate polypectomy technique, 5 = Adjusts / stabilises scope position, 6 = Checks all polypectomy equipment available, 7 = Checks snare closure prior to introduction into the scope, 8 = Clear instructions to and utilisation of endoscopy staff, 9 = Checks diathermy settings are appropriate, 10 = Photo-documents pre and post polypectomy, 11 = Applies prophylactic haemostatic measures if appropriate, 12 = Selects appropriate sized snare, 13 = Directs snare accurately over polyp head, 14 = Correctly selects en-bloc/ piecemeal removal depending on size, 15 = Advances snare sheath towards stalk as snare closed, 16 = Places snare at appropriate position on, 17 = Mobilises polyp to ensure appropriate amount of tissue trapped, 18 = Applies appropriate degree of diathermy, 19 = Examines remnant stalk/ polyp base, 20 = Identifies and appropriately treats residual polyp, 21 = Identifies bleeding and performs adequate endoscopic haemostasis, 22 = Retrieves, or attempts retrieval of polyp, 23 = Checks for polyp retrieval, 24 = Places tattoo completely, where appropriate, 25 = Overall competency at Polypectomy

Participant	DOPyS parameter assessed																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	3	3	4	4	3	3	0	3	3	4	3	3	3	3	3	3	3	3	4	0	0	4	3	0	1
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	4	4	3	4	4	3	3	3	3	3
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	4	4	3	3	4	4	4	3	3	4	0	3	3	0	3	2	3	3	3	0	0	3	3	3	3
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix 4o: DOPyS scores from the **fourth** "real-life" polyp performed following simulator polypectomy training. (Grey-shaded Area representing Group B, unshaded Group A).

ID	Case	Start	Total Time	% of the mucosa examined	Time to reach caecum:	Efficiency of screening: %
1	1.5	1st	00:22:29	93	00:06:11	27
1	1.5	2nd	00:17:47	91	00:04:41	48
1	1.5	3rd	00:14:44	91	00:04:06	65
1	1.6	1st	00:07:56	46	NA	NA
1	1.6	2rd	00:10:25	37	NA	NA
1	1.6	3rd	00:14:34	77	00:04:34	57
1	1.7	1st	00:21:30	90	00:06:01	30
1	1.7	2nd	00:20:24	92	00:05:23	35
1	1.7	3rd	00:18:11	91	00:06:16	46
1	2.1	1st	00:16:29	85	00:02:23	53
1	2.1	2nd	00:15:16	86	00:02:25	60
1	2.1	3rd	00:10:49	83	00:01:59	80
5	1.5	1st	00:29:29	49	NA	NA
5	1.5	2nd	00:15:11	88	00:04:19	NA
5	1.5	3rd	00:10:32	88	00:06:55	NA
5	1.6	1st	00:12:29	81	00:06:49	NA
5	1.6	2nd	00:13:57	80	00:03:23	62
5	1.6	3rd	00:07:44	76	00:03:02	81
5	1.7	1st	00:13:01	88	00:08:25	73
5	1.7	2nd	00:11:09	93	00:06:21	84
5	1.7	3rd	00:09:17	90	00:05:30	91
5	2.1	1st	00:06:45	79	00:02:19	86
5	2.1	2nd	00:08:47	79	00:02:35	83
5	2.1	3rd	00:08:32	80	00:02:06	84

ID	Case	Start	Total Time	% of the mucosa examined	Time to reach caecum:	Efficiency of screening: %
11	1.5	1st	00:10:27	89	00:04:19	86
11	1.5	2nd	00:10:29	91	00:05:37	87
11	1.5	3rd	00:09:14	88	00:03:43	90
11	1.6	1st	00:09:59	79	00:04:06	79
11	1.6	2nd	00:06:56	75	00:02:51	80
11	1.6	3rd	00:09:08	77	00:04:00	79
11	1.7	1st	00:09:46	83	00:04:43	84
11	1.7	2nd	00:10:06	88	00:04:56	87
11	1.7	3rd	00:08:38	88	00:03:40	91
11	2.1	1st	00:08:25	79	00:02:05	83
11	2.1	2nd	00:07:11	81	00:01:49	88
11	2.1	3rd	00:08:26	81	00:01:55	86
13	1.5	1st	00:17:20	90	00:11:06	50
13	1.5	2nd	00:12:42	86	00:07:42	74
13	1.5	3rd	00:04:52	33	NA	NA
13	1.6	1st	00:08:19	68	00:03:55	67
13	1.6	2nd	00:07:51	71	00:03:42	73
13	1.6	3rd	00:05:56	43	NA	NA
13	1.7	1st	00:13:12	85	00:06:20	71
13	1.7	2nd	00:11:43	90	00:06:52	81
13	1.7	3rd	00:14:53	90	00:09:55	64
13	2.1	1st	00:07:58	74	00:02:01	77
13	2.1	2nd	00:07:38	75	00:02:29	79
13	2.1	3rd	00:03:08	18	NA	NA

ID	Case	Start	Total Time	% of the mucosa examined	Time to reach caecum:	Efficiency of screening: %
14	1.5	1st	00:18:25	38	NA	NA
14	1.5	2nd	00:11:54	37	NA	NA
14	1.5	3rd	00:13:23	86	00:06:35	70
14	1.6	1st	00:09:40	44	NA	NA
14	1.6	2nd	00:09:44	50	NA	NA
14	1.6	3rd	00:07:06	45	NA	NA
14	1.7	1st	00:11:11	41	NA	NA
14	1.7	2nd	00:15:42	40	NA	NA
14	1.7	3rd	00:13:31	38	NA	NA
14	2.1	1st	00:05:21	23	NA	NA
14	2.1	2nd	00:03:29	19	NA	NA
14	2.1	3rd	00:11:31	81	00:03:44	75
15	1.5	1st	00:12:47	89	00:04:27	75
15	1.5	2nd	00:20:28	90	00:06:02	34
15	1.5	3rd	00:10:38	82	00:04:58	80
15	1.6	1st	00:17:49	83	00:03:31	45
15	1.6	2nd	00:07:30	75	00:02:43	80
15	1.6	3rd	00:11:57	77	00:02:29	69
15	1.7	1st	00:11:34	90	00:04:15	81
15	1.7	2nd	00:12:24	92	00:06:18	78
15	1.7	3rd	00:08:45	79	00:03:05	82
15	2.1	1st	00:09:57	84	00:02:29	84
15	2.1	2nd	00:09:42	82	00:02:21	83
15	2.1	3rd	00:05:40	71	00:01:18	76

ID	Case	Start	Total Time	% of the mucosa examined	Time to reach caecum:	Efficiency of screening: %
18	1.5	1st	00:12:10	35	NA	NA
18	1.5	2nd	00:24:05	91	00:06:31	22
18	1.5	3rd	00:19:00	93	00:12:10	42
18	1.6	1st	00:13:35	55	NA	NA
18	1.6	2nd	00:13:33	81	00:05:01	65
18	1.6	3rd	00:10:15	79	00:04:24	77
18	1.7	1st	00:09:03	40	NA	NA
18	1.7	2nd	00:12:18	88	00:06:04	77
18	1.7	3rd	00:09:49	89	00:04:19	88
18	2.1	1st	00:04:20	18	NA	NA
18	2.1	2nd	00:10:47	82	00:02:31	78
18	2.1	3rd	00:09:07	79	00:02:58	81
19	1.5	1st	00:12:03	87	00:06:40	78
19	1.5	2nd	00:00:00	0	NA	NA
19	1.5	3rd	00:13:43	65	00:08:23	48
19	1.6	1st	00:08:50	74	00:03:52	76
19	1.6	2nd	00:09:16	70	00:04:30	69
19	1.6	3rd	00:03:30	40	NA	NA
19	1.7	1st	00:07:54	85	00:04:14	91
19	1.7	2nd	00:10:09	73	00:04:29	70
19	1.7	3rd	00:08:11	75	00:03:27	78
19	2.1	1st	00:06:20	80	00:01:57	88
19	2.1	2nd	00:43:22	17	NA	NA
19	2.1	3rd	00:06:45	74	00:02:28	80

Appendix 4p: Results from the Simbionix GI mentor II

Baseline DOPyS parameters	VR group (median IQR)	Ex-vivo group (median IQR)	p value
Attempts to achieve optimal polyp position	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.54
Optimises view by aspiration / insufflation / wash	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.54
Determines the full extent of the lesion	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.74
Uses appropriate polypectomy technique	0.00 (0.00-0.75)	0.00 (0.00-1.75)	0.37
Adjusts / stabilises scope position	0.00 (0.00-0.75)	0.00 (0.00-1.63)	0.37
Checks all polypectomy equipment available	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Checks snare closure prior to introduction into the scope	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Clear instructions to and utilisation of endoscopy staff	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Checks diathermy settings are appropriate	0.00 (0.00-0.38)	0.00 (0.00-1.50)	0.42
Photo-documents pre and post polypectomy	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Applies prophylactic haemostatic measures if appropriate	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Selects appropriate sized snare	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Directs snare accurately over polyp head	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Correctly selects en-bloc/ piecemeal removal depending on size	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Advances snare sheath towards stalk as snare closed	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Places snare at appropriate position on	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Mobilises polyp to ensure appropriate amount of tissue trapped	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.48
Applies appropriate degree of diathermy	0.00 (0.00-0.75)	0.00 (0.00-1.50)	0.42
Examines remnant stalk/ polyp base	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.74
Identifies and appropriately treats residual polyp	0.00 (0.00-0.75)	0.00 (0.00-0.50)	0.89
Identifies bleeding and performs adequate endoscopic haemostasis	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.82
Retrieves, or attempts retrieval of polyp	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.74
Checks for polyp retrieval	0.00 (0.00-0.75)	0.00 (0.00-1.13)	0.74
.Places tattoo completely, where appropriate	0.00 (0.00-0.75)	0.00 (0.00-1.38)	0.89
Overall competency at Polypectomy	0.00 (0.00-0.75)	0.00 (0.00-0.75)	0.96

Appendix 4q**:** Baseline DOPyS scores (Median scores with IQR). Mann-Whitney *U* test used to detect differences between the groups

Appendix 5:

RoTA: a novel objective performance indicator in therapeutic colonoscopy

ID	Assessor 1 Total angulation	Assessor 2 Total angulation	Average Total angulation	Assessor 1 Total Rotation	Assessor 2 Total Rotation	Average Total Rotation	Time taken	RoTA	DOPyS
1	88.00	86.00	87.00	10.00	10.00	10.00	684.00	.10	.50
2	145.00	134.00	139.50	18.00	18.00	18.00	883.00	.10	1.00
3	96.00	119.00	107.50	47.00	45.00	46.00	512.00	.40	1.00
4	115.00	84.00	99.50	121.00	102.00	111.50	712.00	1.10	1.00
5	66.00	75.00	70.50	56.00	52.00	54.00	417.00	1.20	1.50
6	57.00	59.00	58.00	82.00	100.00	91.00	1343.00	1.60	1.50
7	48.00	45.00	46.50	81.00	70.00	75.50	217.00	1.60	2.00
8	138.00	101.00	119.50	128.00	124.00	126.00	489.00	1.00	2.00
9	117.00	115.00	116.00	162.00	141.00	151.50	423.00	1.30	2.00
10	84.00	79.00	81.50	29.00	31.00	30.00	248.00	.40	2.50
11	57.00	49.00	53.00	41.00	32.00	36.50	191.00	.70	2.50
12	58.00	54.00	56.00	22.00	24.00	23.00	272.00	.40	2.50
13	43.00	48.00	45.50	12.00	12.00	12.00	249.00	.30	3.00
14	93.00	78.00	85.50	37.00	42.00	39.50	324.00	.50	3.00
15	183.00	169.00	176.00	83.00	93.00	88.00	255.00	.50	3.00
16	132.00	113.00	122.50	29.00	27.00	28.00	274.00	.20	3.00
17	47.00	38.00	42.50	5.00	5.00	5.00	198.00	.10	3.00
18	46.00	44.00	45.00	4.00	4.00	4.00	168.00	.10	3.00
19	232.00	212.00	222.00	116.00	127.00	121.50	184.00	.50	3.00
20	171.00	186.00	178.50	56.00	48.00	52.00	330.00	.30	3.00
21	73.00	75.00	74.00	24.00	24.00	24.00	242.00	.30	3.50
22	65.00	71.00	68.00	19.00	19.00	19.00	244.00	.30	3.50
23	79.00	77.00	78.00	57.00	66.00	61.50	262.00	.80	4.00

Appendix 5a: Data set for rotational and angulation hand movements

Appendix 6:

Published work

Peer-reviewed publications resulting from thesis work

ANSELL J, Hawkes N, Leicester R, Dolwani S, Torkington J, Warren N

The WIMAT colonoscopy suitcase: a novel polypectomy trainer

Colorectal disease **15** 2012 PP 217-223

ANSELL J, Mason J, Warren N, Dolwani S, Donnelly P, Torkington J

A systematic review of validity testing in colonoscopy simulation

Surgical Endoscopy **26** (11) 2012 PP 3040-52

Mason J, **ANSELL J,** Warren N, Torkington J

Motion Analysis as an assessment tool: A systematic review.

Surgical Endoscopy **27** (5) 2013 PP 1468-77

ANSELL J, Hurley J, Arnaoutakis K, Horwood J, Warren N, Torkington J

Establishing the Construct and Concurrent validity of the WIMAT colonoscopy suitcase

Gastrointest Endoscopy **79** (3) 2014 PP 490-7

ANSELL J, Hurley J, Arnaoutakis K, Horwood J, Warren N, Torkington J

Can trainees accurately self-assess their competency in polypectomy?

American Journal of Surgery **207**(1) 2014 PP 32-8.

Academic prize awarded from thesis work

The WIMAT colonoscopy suitcase: a novel polypectomy simulator

Steve Lovell Audio-Visual Prize, Association of Surgeons Great Britain and Ireland 2013

Presentations to learned societies (International)

ANSELL J, Warren N, Donnelly P, Hawkes N, Dolwani S, Torkington J

Systematic review of validity testing in colonoscopy simulation

23rd Colorectal Disease Symposium, Cleveland Clinic Florida, Feb 2012

ANSELL J, Warren N, Donnelly P, Hawkes N, Dolwani S, Torkington J

Systematic review of validity testing in colonoscopy simulation

European Association of Endoscopic Surgery (EAES) Brussels, June 2012

ANSELL J, Hawkes N, Leicester R, Dolwani S, Torkington J, Warren N

The WIMAT colonoscopy suitcase: a novel polypectomy trainer

European Association of Endoscopic Surgery (EAES) Brussels, June 2012

ANSELL J, Hawkes N, Leicester R, Dolwani S, Torkington J, Warren N

The WIMAT colonoscopy suitcase: a novel polypectomy trainer

Society of American Gastrointestinal & Endoscopic Surgery (SAGES) Baltimore, 2013

ANSELL J, Hurley J, Horwood J, Rizan C, Warren N, Torkington J

Can colonoscopists self-assess performance using the WIMAT colonoscopy suitcase?

Society of American Gastrointestinal & Endoscopic Surgery (SAGES) Baltimore, 2013

ANSELL J, Hurley J, Horwood J, Rizan C, Warren N, Torkington J

The WIMAT colonoscopy suitcase is a valid simulator for polypectomy training

Society of American Gastrointestinal & Endoscopic Surgery (SAGES) Baltimore, 2013

ANSELL J, Hurley J, Horwood J, Glasbey J, Warren N, Torkington J

ROtation To Angulation (ROTA): The development of an objective scoring system for colonoscopic polypectomy

Tripartite Colorectal Meeting , Birmingham 2014

Presentations to learned societies (National)

ANSELL J, Warren N, Donnelly P, Hawkes N, Dolwani S, Torkington

Systematic review of validity testing in colonoscopy simulation

Association of Surgeons in Training (ASiT) Cardiff, March 2012

ANSELL J, Warren N, Donnelly P, Hawkes N, Dolwani S, Torkington

Systematic review of validity testing in colonoscopy simulation

Association of Surgeons of Great Britain and Ireland (ASGBI) Liverpool May 2012

Mason J, **ANSELL J,** Warren N, Donnelly P, Torkington J

Motion Analysis as an assessment tool: A systematic review

Association of Surgeons in Training (ASiT) Cardiff, March 2012

Mason J, **ANSELL J,** Warren N, Donnelly P, Torkington J

Motion Analysis as an assessment tool: A systematic review

Association of Surgeons of Great Britain and Ireland (ASGBI) Liverpool May 2012

ANSELL J, Hurley J, Horwood J, Rizan C, Warren N, Torkington J

The WIMAT colonoscopy suitcase, a valid simulator for polypectomy training

Association of Surgeons of Great Britain and Ireland (ASGBI) Glasgow May 2013

ANSELL J, Hurley J, Horwood J, Rizan C, Warren N, Torkington J

Can colonoscopists self-assess performance using the WIMAT colonoscopy suitcase?

Association of Surgeons of Great Britain and Ireland (ASGBI) Glasgow May 2013

Surg Endosc (2012) 26:3040–3052
DOI 10.1007/s00464-012-2332-2

REVIEW

Systematic review of validity testing in colonoscopy simulation

James Ansell · John Mason · Neil Warren · Peter Donnelly · Neil Hawkes · Sunil Dolwani · Jared Torkington

Received: 17 November 2011 / Accepted: 25 April 2012 / Published online: 31 May 2012

Abstract
Background Simulation is a useful adjunct to skills-based training. It potentially avoids risk to patients during training and development of basic interventional techniques. This may be of particular relevance in colonoscopy where the learning curve can be long. Several endoscopic devices exist that simulate colonoscopy for training purposes. This study was designed to review the evidence for the validity of these simulators.

Methods MEDLINE (1947 to present), PubMed, Embase classic + Embase, the *meta*Register of Controlled Trials, and the Education Resources Information Center (ERIC) were searched for studies validating colonoscopy simulators. For each study, we recorded the type of simulator used, the tasks assessed, the endpoints reported, and the type of validity measured. Common endpoints between studies were compared, and the evidence was graded.

Results Thirteen studies met the inclusion criteria. Construct validity was reported in five (41.7 %) studies for the Accutouch HT Immersion (cases 1, 3, and 4), four studies (33.3 %) for the GI mentor II (Simbionix) (Modules 1.1, 1.3, 1.7, 2.1, and 5), two studies (16.7 %) for the Olympus Endo Ts-1 2nd Generation, and one study for the Endo X bovine model. Face validity was reported for the Accutouch HT Immersion, the Olympus 2nd Generation, and the KAIST-Ewha. Content validity was reported for the all simulators, excluding the KAIST-Ewha. The only report of criterion validity was for the Endo X bovine model.

Conclusion Evidence exists to support the face, content, and construct validity of several virtual reality colonoscopy simulators for specific diagnostic and therapeutic modules with selected endpoints. One study demonstrates content, construct, and criterion validity for an ex vivo animal platform. Further work is needed to demonstrate the criterion validity of all devices.

Keywords Validity · Simulation · Endoscopy · Education

J. Ansell (✉)
Royal College of Surgeons Clinical Research Fellow in Simulation, Welsh Institute for Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, Wales CF14 4UJ, UK
e-mail: ansellj@cf.ac.uk

J. Mason
Cardiff University School of Medicine, Heath Park, Cardiff, Wales CF14 4UJ, UK

N. Warren
Welsh Institute for Minimal Access Therapy, Heath Park, Cardiff, Wales CF14 4UJ, UK

P. Donnelly
Wales Deanery, Neuadd Meirionnydd, Heath Park, Cardiff CF14 4YS, UK

N. Hawkes
Royal Glamorgan Hospital, Llantrisant, Cardiff, Wales CF72 8XR, UK

S. Dolwani
University Hospital Llandough, Penlan Road, Penarth, Cardiff, Wales CF64 2XX, UK

J. Torkington
University Hospital of Wales, Heath Park, Cardiff, Wales CF14 4XN, UK

There is concern amongst some medical educators that reduced training hours may ultimately have a detrimental impact on clinical competency [1–4]. Therefore, a paradigm shift has begun in most craft specialities to streamline training. Increasing emphasis is being placed on focussing

objectives to the individual needs of the trainee. Simulation offers a useful adjunct to conventional methods of training. It allows the participant to gain the skills needed to progress along their learning curve in a safe environment [5–7]. In addition, there is evidence that simulators may play a role in formative assessment.

In particular, the field of endoscopy has embraced simulation, providing training opportunities outside the clinical setting [8]. In the United Kingdom, this is approved by the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) [9]. Several types of colonoscopy simulators are currently available with the ability to teach diagnostic and therapeutic interventions [10]. (A full description of each of these is outlined in Table 1).

Validation and reliability

Validation is the extent to which an instrument measures what it was designed to measure [11, 12]. There are several forms (Table 2). The first stage of any validation process is to establish a "construct" [13]. This defines exactly what needs to be examined by a new training tool [13]. In colonoscopy, for example, a robust "construct" should clarify whether the simulator is being validated to assess or to teach its users [13]. Once this is established, the process of validation should provide evidence to support each facet. The more aspects of validity that are proven, the stronger the overall argument is [13]. It is vital that simulator validity is fully tested to ensure that its full effectiveness and clinical application is established.

Reliability is the aptitude of an instrument to discriminate consistently between performance across evaluators or over time [12]. It is measured on a scale of 0 to 1, with 0 being totally unreliable and 1 being completely reliable [14, 15]. It is generally agreed that an appropriate score of reliability for a test is $\geq$0.8 [15].

The purpose of this systematic review was to evaluate the evidence for validity testing of colonoscopy simulators and, in turn, grade the strength of this research. This may suggest ways in which colonoscopy simulators should be utilized in medical training.

Materials and methods

Data sources, search strategy, and data extraction

A systematic review of published work was conducted according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines [16]. We used the following sources to search for studies reporting on the validity of colonoscopy simulators: MEDLINE (1947 to present), PubMed (from 1966 to May 2010), Embase

Table 1 Summary of endoscopic simulators for performing diagnostic and therapeutic colonoscopy

Simulator	Examples	Description
1. Mechanical	Adam Rouilly simulator	Silicone rubber colonic model for basic colonoscopy procedure.
	Koken Model l-B	Lumen of simulated bowel shaped to replicate different areas of the colon
	Chamberlain Group LLC [50]	Plastic colon, mounted in rigid foam. Has the ability to be fitted with replaceable colonic stricture and polyps. This group also makes a straight colon section to train polypectomy and stenting [50]
2 Composite animal	Endo X Trainer™ (Medical Innovations)	Portable plastic tray system, can lay animal tissues within the tray and can perform a variety of procedures without or without simulated bleeding [10, 30]
	Colonoscopy suitcase (WIMAT)	Portable colonoscopy trainer which can simulate removal of sessile and pedunculated polyps with/without the capacity for bleeding and diathermy [52]
3. Virtual reality	Accutouch HT (CAE Previously Immersion Medical)	Trolley-mounted, computerized device with a flat-screen display on a movable arm. A model endoscope is provided with the system. Several modules are available range of endoscopic procedures and pathology. It simulates patient vital signs and responses to administration of sedation and to pain [10]
	GI mentor II (Simbionix)	Contains library with more than 120 tasks [51]. Can measure end points including procedure time, visualization of bowel, mechanical pressures on bowel [51]. Has modules for degrees of difficulties and dealing with pathology
	KAIST-Ewha	Manages training scenarios with varying degrees of difficulty, measures multiple parameters, including time taken, exertion force, tip motion, number of red outs [29]
	EndoTSI (2nd Gen) (Olympus)	For training and assessment of colonoscopy skills. Simulates multiple matrices, including shaft looping, tip contact, variable shaft stiffness, application of abdominal pressure, and movement of patient [22]
4. Live animal models	Porcine Model	A realistic platform with haptic feedback identical to human tissue. These are expensive with ethical concerns and can demonstrate anatomical variation

Table 2 Summary of validity types

Type of validation	Definition	Method of examination
Face	The rational expectation that an association between two things exists [13]	Questionnaire to non-expert users
Content	The extent to which a measure reflects the trait or domain it purports to measure [12]	Questionnaire to expert users
	For training tool: need to data prove haptic and visual realism [13]	
	For assessment tool: review by experts of the skill domains being tested [13]	
Construct/contrast	An agreement between a theoretical concept and a specific tool or procedure [12] (experienced surgeons should score higher on its assessment parameters than juniors)	Measuring relevant parameters for defined groups of variable expertise
Criterion (*predictive or concurrent*)	*Predictive validity*: The ability of a tool to predict future performance [12]	Correlation between test score and future performance ratings
	Concurrent validity: The correlation between assessment tool and the "gold standard" [12]	Comparison with patient based data

classic + Embase (1947–2011 week 32), the *meta*Register of Controlled Trials, and the Education Resources Information Center (ERIC). We used three different domains of exploded MeSH-terms and key words combined by "AND." Within each domain, the terms were combined with "OR." The first domain contained the terms for endoscopy, the second for simulation, and the third for validation. Two investigators reviewed titles and abstracts resulting from the search separately using a predefined data extraction form. Articles were retrieved when judged to meet potentially the inclusion criteria. Investigators then independently applied the inclusion criteria to retrieved articles. Those clearly unrelated to the topic and any duplicates were excluded. Any differences were referred to a third party for final analysis. Articles for cross-referencing at this stage also were included for final evaluation. The last search date was September 15, 2011.

Inclusion and exclusion criteria

All studies validating colonoscopy simulators for assessment or training were included in the review. Included studies needed to contain sufficient details of the simulation model used and type of validity measured. Foreign language studies were included and converted to English by using a computer translation program [17]. Reviews, congress abstracts, and studies that validated tools of assessment and not the simulator itself were excluded. Studies detailing endoscopic methods other than colonoscopy also were excluded.

Data extraction, outcome measures, and analysis

Included studies were rated by a guideline based on the Centre for Evidence-Based Medicine (CEBM), levels of evidence (Tables 3, 4) [18]. For each paper, we recorded the type of simulator used, the task being assessed, the endpoints of the study, and the type of validity measured.

Table 3 Summary of evidence levels [18, 31]

Level	Diagnosis
Ia	Systematic reviews (metaanalysis) containing at least some trials of Level Ib evidence, in which results of separate, independently conducted trials are consistent
Ib	RCT of good quality and of adequate sample size (power calculation)
IIa	RCT of reasonable quality and/or of inadequate sample size
IIb	Nonrandomized trials, comparative research parallel cohort
IIc	Nonrandomized trial, comparative research (historical cohort, literature controls)
III	Nonrandomized, noncomparative trials, descriptive research
IV	Expert opinions, including the opinion of work group members

Table 4 Summary of recommendation levels [18, 31]

Level of recommendation	Criteria
I	One systematic review (Ia) or at least two independently conducted research projects classified as Ib
II	At least two independently conducted research projects classified as Levels IIa or IIb, within concordance
III	One independently conducted research project Level IIb, or at least two Level III trials, within concordance
IV	One trial at Level III or multiple expert opinions, including the opinion of work group members

Common endpoints between papers were compared when statistically significant results were reported. The principle summary statistic was the difference in means. It was judged that the data were not suitable for statistical pooling due to the heterogeneity of study design.

Results

The primary search identified 1,141 studies. After duplicates were removed, 739 titles and abstracts were screened for relevance. From this, 678 records were excluded, leaving 61 full text articles for review. After a review of the full text, 53 records were excluded and 5 cross-referenced, leaving a total of 13 papers for inclusion (Fig. 1; Table 5). Twelve (92.3 %) of these articles validated VR simulators, which included four (33.3 %) GI mentor II (Simbionix^{TM}) studies [8, 19–21] (Modules 1.1, 1.3, 1.7, 2.1, and 5), two (16.6 %) Olympus Endo Ts-1 2nd Generation studies [22, 23], five (41.6 %) Accutouch HT Immersion studies [24–28] (cases 1, 3 and 4), and one (8.33 %) KAIST-Ewha Simulator study [29]. One novel study looked at a composite model using bovine intestine and the Endo X colonoscopy platform [30]. No randomized, controlled trials have addressed the issue of validation. The majority of evidence was classed as IIb or III (Table 3). The levels of recommendation were calculated according to the overall number trials for each simulator (Table 4).

VR simulators

Face, content, and criterion validity

Face validity was reported in two studies for the Olympus Endo Ts-1 2nd Generation [22, 23] and in two studies for the Accutouch HT Immersion [26, 27] (Level III recommendation). One study reported face validity for the KAIST-Ewha [29] (Level IV recommendation). Content validity (obtained from experts completing realism surveys) was

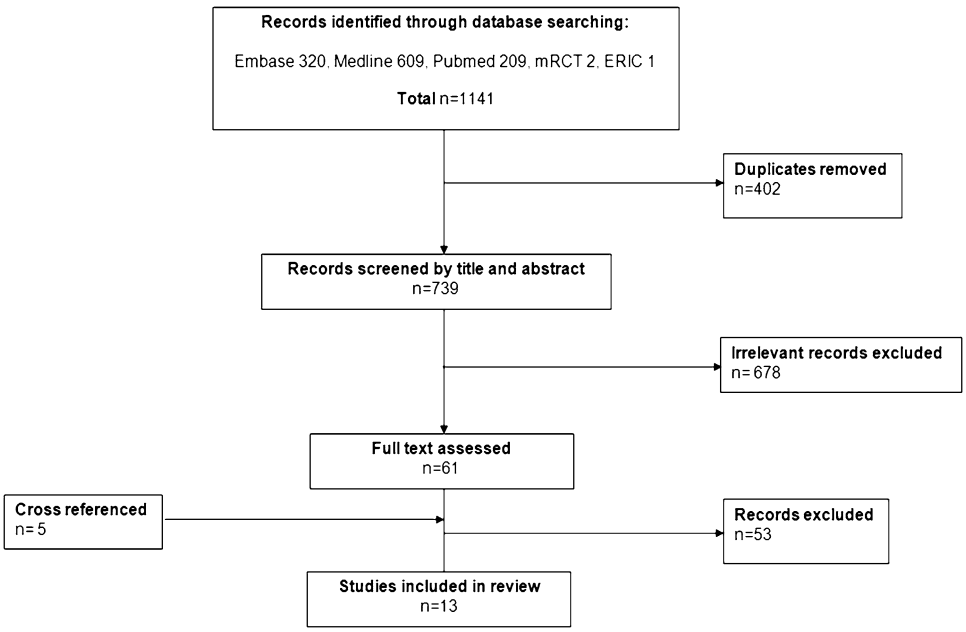


Fig. 1 Flow *diagram* for search strategy

Table 5 Summary of included studies

Study	Simulator tested	Simulator class
Datta et al. [24]	Accutouch HT Immersion	Virtual reality
MacDonald et al. [25]	Accutouch HT Immersion	Virtual reality
Mahmood and Darzi [26]	Accutouch HT Immersion	Virtual reality
Sedlack and Kolars [27]	Accutouch HT Immersion	Virtual reality
Moorthy et al. [28]	Accutouch HT Immersion	Virtual reality
Felsher et al. [20]	Simbionix GI mentor II	Virtual reality
Grantcharov et al. [21]	Simbionix GI mentor II	Virtual reality
Sedlack et al. [30]	Novel bovine model	Composite
Koch et al. [23]	Olympus 2nd Generation	Virtual reality
Koch et al. [8]	Simbionix GI mentor II	Virtual reality
Woo et al. [29]	KAIST-Ewha	Virtual reality
Haycock et al. [22]	Olympus 2nd Generation	Virtual reality
Fayez et al. [19]	Simbionix GI mentor II	Virtual reality

Table 6 Summary of results for VR content validation

Simulator	Study	Expert questionnaire	Measured parameters	Likert value
Accutouch HT Immersion	Sedlack and Kolars [27]	10-Point scale	Realism of controls	7.9
			Visual graphics	7.0
			Force/feel	6.5
			Insufflations and suction	5.7
			Scope controls	8.0
			Loop management	6.6
Olympus 2nd Generation	Koch et al. [23]	10-Point scale	Difficulty	72
			Practical setup	6.9
			Endoscope handling	7.6
			Display of endoscope movement	7.1
			Tactile feedback	7.0
			Insertion	7.1
	Haycock et al. [22]	10-Point scale	Appearance	5.8
			Movement	6.4
			Force feedback	6.6
			Looping	6.6
			Loop resolution	6.8
GI mentor II Simbionix	Koch et al. [8]	4-Point scale	Overall realism	2.95
			Anatomical representation	2.58
			Simulator setup	3.14
			Endoscopic control	3.21
			Haptic feedback	2.57
Endo X Animal Model	Sedlack et al. [30]	7-Point scale	Mucosal realism	6.0
			Endoscopic view	6.0
			Degree of paradoxical motion	6.0
			Resistance	5.0
			Overall fidelity	5.0

reported in one study each for the GI mentor II (Simbionix^{TM}) [8] and the Accutouch HT Immersion [27] (Level IV recommendation; Table 6). Content validity also was reported in two studies for the Olympus Endo Ts-1 2nd Generation [22, 23] (Level III recommendation; Table 6). There were no reports of criterion validity for any VR simulator.

Construct validity

Construct validity was reported in 11 (91.6 %) of the 12 VR studies included [8, 19–28], but the methodology varied between them. In all studies that examined construct validity, participants were assigned to groups according to the previous number of colonoscopies they had performed. Each study had different criteria for this (Table 7). Participants were asked to complete a variety of modules/cases assessing navigation, diagnostics, or therapeutics. The nature of these tasks also varied (Table 7). Several common endpoints were comparable between studies that validated the same simulator (Table 8). The results of these common endpoints are as follows:

GI mentor II (Simbionix) Four studies confirmed construct validity of the GI mentor II (Simbionix) simulator [8, 19–21] for the following common endpoints: procedure time [19, 21], efficiency [19–21], loop formation [19–21] (Level II recommendation), cecal intubation time [20], polypectomy rate [20], and %mucosa visualized [21] (Level III recommendation; Modules 1.1, 1.3, 1.7, 2.1, and 5). One paper did not reference the module that they were assessing [21]. One study [8] failed to demonstrate construct validity for any endpoints when comparing intermediate, experienced, and expert users.

Accutouch HT Immersion Five studies demonstrated the construct validity of the Accutouch simulator [24–28] for the following common endpoints: total procedure time [24–27], %mucosa visualized [24, 26] (Level II recommendation), and polypectomy rate [25], efficiency [24],

Table 7 Summary of methodology

Study (simulator)	*n*	Groups (no. of previous colonoscopies)	Validation	Modules/cases	Task	Endpoints
Fayez et al. [19] (GI mentor II)	20	N (<5 scopes) $n = 12$ E (>50 scopes) $n = 8$	Construct	Module 1 (cases 1 and 7) Module 2 (case 1)	Navigation Polypectomy	Time, red out, excessive pressure, %mucosa, %clear view, %pain, looping, %efficiency
Felsher et al. [20] (GI mentor II)	75	Cohort 1 ($n = 37$) N: $n = 14$; E: $n = 23$ Cohort 2 ($n = 38$) N: $n = 13$; E: $n = 25$	Construct	Co 1, Modules 1 and 5 Co 2, Module 2 (case 1)	Navigation Bx Polypectomy	Time to cecum, %mucosa, polypectomy rate %biopsy rate, %time to clear view, time in pain, efficiency ratio
Grantcharov et al. [21] (GI mentor II)	28	N (0 scopes) $n = 10$ I (<50 scopes) $n = 10$ E (>200) $n = 8$	Construct	Case 1	Navigation	Time, %mucosa, efficiency, time with clear view, excessive local pressure, pain, time with pain, loop formation and time with loop
Koch et al. [8] (GI mentor II)	105	N (0 scopes) $n = 35$ I (<200 scopes) $n = 15$ E (200–1,000 scopes) $n = 20$ Expert (EX) (>1,000 scopes) $n = 35$	Construct Content	Module 1 (cases 1 and 3) 1 Hand-eye test Realism survey	Navigation	Time to cecum, %time with clear view, lost view of lumen, excessive local pressure, %time in pain, excessive loop formation
Datta et al. [24] (HT Immersion)	45	N (0 scopes) I (5–50 scopes) E (>200 scopes)	Construct	Case 1 Case 4	Navigation	%Mucosa visualised, time taken, path length red-out and efficiency ratio[a]
MacDonald et al. [25] (HT Immersion)	34	N: $n = 10$ I: $n = 19$ E: $n = 5$	Construct	Three scopes increasing difficulty with pathology to identify	Navigation identify lesion	Time, insertion length, %bowel visualised, pain, red out, air left in bowel, max force, air levels, %cancers/lesions visualized, %perforations.
Mahmood and Darzi [26] (HT Immersion)	25	N: (<10 scopes) $n = 11$ I (11–100 scopes) $n = 7$ E (>101 scopes) $n = 7$	Construct Face	Modules 3 or 4	Navigation	Time, %mucosa, path length, mean incidence of perforation

Table 7 continued

Study (simulator)	*n*	Groups (no. of previous colonoscopies)	Validation	Modules/cases	Task	Endpoints
Sedlack and Kolars [27] (HT Immersion)	22	N (0 scopes) $n = 6$	Content	Cases 3 and 4	Navigation	%Mucosa, time, pain, max depth insertion, air insufflated, air remaining, max force, %complications, %identification of pathology
		I (88–202 scopes) $n = 6$	Construct	Realism survey	Identify lesion	
		E (482–694 scopes) $n = 10$	Face			
Moorthy et al. [28] (HT Immersion)	20	N (1–10 scopes) $n = 7$	Construct	Case 4, analysis of endoscopic/hand view	Navigation	Hand movement, force, endoscopic handing, red out, patient pain, flow of procedure, time taken, depth of scope insertion%mucosa
		I (20–80 scopes) $n = 7$			Hand movement	
		E (>200 scopes) $n = 6$				
Haycock et al. [22] (Olympus)	34	N (0 scopes) $n = 10$	Face	Realism survey	Navigation	Caecal intubation, time taken, excess inflation, variable stiffness use, shaft insertion parameters, loop formation and time with loop
		I (<1,000 scopes) $n = 13$	Construct	Various loop formations[a]	Loop tasks	
		E (>1,000 scopes) $n = 11$	Content			
Koch et al. [23] (Olympus)	49	N (0 scopes) $n = 26$	Face content	Realism survey	Navigation	Dexterity, time to cecum, shaft insertion force, shaft torque, tip section force, max patient pain
		E (>1,000 scopes) $n = 23$	Construct	Sigmoid N loop transverse γ loop		
Sedlack et al. [30] (Bovine model)	39	N (no scopes) $n = 13$	Content	X1 simulated colon	Navigation	Cecal intubation, depth insertion, time, scope length to reach cecum, %mucosa, quality of mucosal examination
		I (100–150) $n = 13$	Construct	Realism survey		
		E (high-volume centers) $n = 13$	Criterion			
Woo et al. [29] (KAIST-Ewha)	5	E (2,500–5,000 scopes) $n = 3$	Face	Realism survey after three colonoscopies	Navigation	Ten items to evaluate realism
		I (1,000 scopes) $n = 2$				

There is wide variation between each study's definition of novices, intermediates, and experience participants. There also are differences in the modules selected for validation between papers. A degree of overlap between study endpoints is exhibited *N* novice, *I* intermediate, *E* experienced, *Ex* expert, *n* number of participants in group

[a] Sigmoid N and moderate transverse loop, sigmoid N moderate transverse loop with low pain tolerance, sigmoid α moderate transverse loop

Table 8 Statistical outcomes for studies detailing construct validation (p values were significant at <0.05)

Sim type	Study	Group/ module	Procedure time	Cecal intubation time	Efficiency/ efficiency ratio	Loop formation	Patient pain	Time with clear view	%Cancers/ lesions visualized	%Mucosa visualized	Polypectomy/ Bx rate (%)
Simbionix GI mentor II	Fayez et al. [19]	N vs. E 1.1	**<0.01**	–	0.08	0.38	0.3	0.61	–	0.13	–
		N vs. E 1.7	**<0.01**	–	**<0.01**	0.93	0.58	0.05	–	**0.01**	No data
		N vs. E 2.1	**0.03**	–	**<0.01**	0.64	0.9	0.11	–	0.07	No data
	Felsher et al. [20]	N vs. E gp 1	–	0.07	0.02	–	0.6	0.05	–	**0.02**	0.03/0.7
		N vs. E gp 2	–	**0.01**	**0.03**	–	0.3	0.9	–	0.30	0.01
	Grantcharov et al. [21]	N vs. L vs. E	**<0.001**	–	**0.001**	**<0.001**	**0.004**	**0.001**	–	**0.001**	–
	Koch et al. [8]	Ex vs. E I 1.1	–	0.962	–	**0.02**	0.077	0.621	–	0.621	–
		Ex vs. I 1.1	–	0.141	–	**0.018**	**0.018**	0.259	–	0.259	–
		I vs. E 1.1	–	0.166	–	0.547	0.385	0.617	–	0.617	–
		I vs. N 1.1	–	**0**	–	0.743	0.07	0.177	–	0.177	–
		Ex vs. E 1.3	–	0.969	–	0.726	0.154	0.297	–	0.297	–
		Ex vs. I 1.3	–	0.326	–	0.09	0.111	0.757	–	0.757	–
		1 vs. E 1.3	–	0.257	–	0.184	0.771	0.394	–	0.394	–
		I vs. N 1.3	–	**0**	–	**0.04**	0.584	0.104	–	0.104	–
Accutouch HT Immersion	Datta et al. [24]	N vs. I vs. E	**<0.001**	–	**<0.001**	–	–	–	–	**<0.001**	–
	MacDonald et al. [25]	E vs. I	>0.05	–	–	–	>0.05	–	>0.05/>0.05	>0.05	–
		I vs. N	>0.05	–	–	–	**<0.001**	–	**<0.001/<0.05**	>0.05	–
		E vs. N	**<0.001**	–	–	–	>0.05	–	**<0.001/<0.05**	>0.05	–
	Mahmood and Darzi [26]	N vs. I vs. E	**0.005**	–	–	–	–	–	–	**0.001**	–
		N vs. I vs. E 3	**0.030**	–	–	–	–	–	–	**0.001**	–
	Sedlack and Kolars [27]	E vs. I	**<0.01**	**<0.01**	–	–	>0.05	–	>0.05	>0.05	>0.05
		I vs. N	**<0.01**	**<0.01**	–	–	>0.05	–	>0.05	>0.05	>0.05
		E vs. N	**<0.01**	**<0.01**	–	–	>0.05	–	>0.05	>0.05	>0.05
	Moorthy et al. [28]	N vs. I vs. E	0.75	–	–	–	–	–	–	0.21	–
Olympus 2nd Generation	Haycock et al. [22]	E vs. I	7.00	7.00	–	7.00[a]	N vs. I vs. E 0.43	–	–	–	–
		I vs. N	**<0.001**	**0.003**	–	**0.05**[a]		–	–	–	–
		E vs. N	**<0.001**	**0.001**	–	**0.02**[a]		–	–	–	–
	Koch et al. [23]	E vs. N	–	**0.001**	–	–	**0.018**	–	–	–	–

Table 8 continued

Sim type	Study	Group/ module	Procedure time	Cecal intubation time	Efficiency/ efficiency ratio	Loop formation	Patient pain	Time with clear view	%Cancers/ lesions visualized	%Mucosa visualized	Polypectomy/ Bx rate (%)
Endo X	Sedlack et al. [30]	E vs. 1	**0.036**	**0.001**	**<0.001**	–	–	–	–	**0.001**	–
		I vs. N	**<0.001**	**0.011**	**<0.001**	–	–	–	–	**0.002**	–
		E vs. N	**<0.001**	**0.011**	**<0.001**	–	–	–	–	**<0.001**	–

Ex expert, *E* experienced, *I* intermediate, *N* novice, – data unavailable or not measured in study

[a] Time with sigmoid loop after intubation

Several studies report statistically significant results (bold) to validate simulator construct for procedural and cecal intubation time. There is sporadic evidence for the construct validity of other VR simulator endpoints

and cecal intubation time [27] (Level III recommendation; cases 1, 3, and 4). Two studies [25, 27] failed to demonstrate construct validity for several endpoints when comparing intermediate, experienced users.

Olympus Endo Ts-1 2nd Generation Two studies [22, 23] reported the construct validation of the Olympus Endo Ts-1 (2nd Generation) simulator for the following common endpoints: cecal intubation, time procedure time, pain and loop formation [22, 23] (Level III recommendation; cases: sigmoid N loop moderate transverse loop ± low pain threshold, sigmoid α loop moderate transverse loop, and sigmoid N loop with γ transverse loop). In one study [22], construct validity was not demonstrated for these endpoints in the expert versus intermediate group. There was no evidence for the construct validity of therapeutic measures in these two studies.

Composite animal-model simulators

Content and criterion validity

One study [30] reported the content validity of the Endo X bovine model (Level IV recommendation). Realism scores for the endoscopic mucosal appearance and degree of paradoxical motion were favourable compared to VR (Table 6). Interrater reliability was confirmed. This study also demonstrated preliminary criterion validity of the simulator for cecal intubation time (Level IV recommendation). They showed that cecal intubation times on the simulator were comparable to their individual patient-based times obtained from a training database.

Construct validation

There is one published study [30] that demonstrated the construct validity of Endo X bovine model for the following endpoints: procedure time, cecal intubation time, and %mucosa visualized (Level III recommendation).

A summary of validity evidence for specific cases/ modules and endpoints along with the grading and level of recommendation can be found in Table 9.

Discussion

Endoscopic training consists of developing cognitive, clinical, and technical skills, the latter being traditionally acquired through mentoring [31, 32]. This results in a potential added risk to patient safety and comfort [33]. With the recent reduction in working hours, an alternative training model has been sought. Simulation in colonoscopy may be well placed to supplement the needs of the modern

Table 9 Level of validity evidence and recommendation for each colonoscopy simulator (*) (**) are used to group together VR endpoints so that evidence levels can be assigned

Sim	Cases/modules/task (construct of study)	Face (evidence level)	Content (evidence level)	Criterion (evidence level)	Construct validity (evidence level)	Recommendation level
GI mentor (Simbionix)	Modules: 1.1, 1.3, 1.7, 2.1, and 5 (training)	No	Yes (3) [8]	No	Yes, for total procedure time (IIb) [19, 21] (*) Yes, for loop formation (IIb) [19–21] (*) Yes, for efficiency (IIb) [19–21] (*) Yes, for cecal intubation time (IIb) [20] (**) Yes, for polypectomy rate (IIb) [20] (**) Yes, for %mucosa visualized (IIb) [21] (**)	Level IV for content Level II for construct for (*) endpoints Level III for construct validity for (**) endpoints
Accutouch HT Immersion	Cases: 1, 3, and 4 (assessment and training)	Yes (3) [26, 27]	Yes, realism (3) [27]	No	Yes, for total procedure time (IIb) [24–27] (*) Yes, for %mucosa visualized (IIb) [24, 26] (*) Yes, for polypectomy rate (IIb) [25] (**) Yes, for efficiency (IIb) [24] (**) Yes, for cecal intubation (IIb) [27] (**)	Level III for face Level IV for content Level II for construct for (*) endpoints Level III for construct for (**) endpoints
Olympus 2nd Generation	1. Sigmoid N and transverse loop ± pain threshold 2. Sigmoid α loop and transverse loop 3. Sigmoid N loop with γ transverse loop (assessment and training)	Yes (3) [22, 23]	Yes, realism (3) [22, 23]	No	Yes. some evidence for cecal intubation time procedure time, pain, and loop formation (IIb) [22, 23] (not for experts vs. intermediates)	Level III for face and content Level II for construct for some endpoints
KAIST-Ewha	Navigation colonoscopy (training)	Yes (3) [29] reliable survey (Cronbach α 0.8)	No	No	No	Level IV for face
Endo X Animal	Navigation colonoscopy (assessment)	No	Yes, realism (3) [30]	Yes, concurrent for intubation times (IIc) [30]	Yes, for total procedure time, cecal intubation time, %mucosa visualized, and efficiency (IIb) [30] (*) Interrater reliability of process proved	Level IV for content and criterion Level III for construct for (*) endpoints

trainee. It is therefore important that the evidence for simulator validity is examined.

The use of simulators for the acquisition of practical skills is well described [34–36]. In the past, there has been a focus toward laparoscopic surgery. This has led to the publication of a Cochrane Collaboration Review [37]. This advocates the use of laparoscopic simulation for supplementing standard training methods [37]. When considering endoscopic simulation, the literature base is not as widespread. Looking particularly at endoscopic simulator validation, there is a weight of evidence that focuses on upper GI simulators and therapeutics [38–40]. Where there is validity research of lower GI endoscopy simulation, the focus is on VR. This evidence has been previously reported by Carter et al. [31], who published a document entitled, *Consensus guidelines for the validation of virtual reality surgical simulators*. They highlight a lack of published validation studies with vigorous experimental methodology for endoscopy [13]. They concluded: Level II recommendation for the overall validity of the Accutouch HT Immersion, diagnostic cases 1, 3, and 4 (for total time, %mucosa seen, and path length) with no evidence for therapeutics. In addition, at that time, there was scant evidence to support the validity of the Simbionix GI mentor [13].

Since this review, the evidence base for these simulators has progressed. Further validity studies have been published, and new endoscopic simulators have been introduced. From this systematic review, we can now report new validation evidence for the Accutouch HT Immersion, the GI mentor II (Simbionix) simulator, the Olympus Endo Ts-1 (2nd Generation), the KAIST-Ewha, and the Endo X bovine model. This includes evidence for VR polypectomy models. Our review shows that the majority of evidence at present is for construct validity of VR simulators. There is some evidence to support face and content validity of VR simulators. There is no reference to face validity for the GI mentor II (Simbionix) or the Endo X bovine model in the literature. However, it is likely, by definition, that these do possess face validity because their content and construct validity are confirmed. There is no evidence demonstrating criterion validity for VR. Furthermore, there is only one paper that reports content, construct, and criterion validity of an animal tissue model. The reason for the focus toward VR in the literature can be explained in two ways. First, the endpoints for each simulation are recorded by the instruments' computer system. This standardizes the assessment and makes data collection a relatively undemanding process. Second, VR modules/cases can create a range of procedures of varying difficulty with the click of a button. This avoids the need for a time consuming "set-up" process that can accompany animal tissue validity testing.

Although the currently available computer simulation scenarios are very reproducible, some argue that they lack the complexity and fidelity to be useful in any meaningful way [30]. Several studies show that VR simulators often are unable to distinguish consistently between intermediate level trainees and experts across all endpoints (Table 8) [8, 22, 25]. This strengthens the argument that VR at present may only be useful for teaching basic endoscopic skills to the novice trainee. It may be that the level of difficulty needs to be increased to make the distinction between more experienced groups. Indeed, this is a view shared by several papers [41–43]. A systematic review by Sturm et al. [44–48] demonstrated that VR simulation colonoscopy training provided participants with an advantage over their untrained colleagues when performing the task for real. This was particularly noticeably at the initial stages of learning, after which there was no difference [45]. An alternative option is the composite animal model. Sedlack et al. [30] has shown the ability of the bovine model to discriminate between senior trainees and experts for several endpoints. This coupled with favorable realism scores and difficulty ratings creates an area of interest for future research.

This review highlights that included studies reported a range of simulator endpoints. These were used as an indicator of performance to confirm construct validation. The papers included in this review often differed in the endpoints that they used, although there was a degree of overlap (Tables 7, 8). Several studies reported endpoints that were unable to distinguish between levels of experience in a statistically significant way. This questions the construct validity of these endpoints. Several researchers in the past have suggested that VR endpoints are with few exceptions invalid as meaningful metrics [27, 49]. From this review, we report that the most valid training and assessment endpoints across all studies are: total procedure time, cecal intubation time, efficiency, and %mucosa visualized. These could be considered as the most reliable indicators of performance in future work. Trainees using VR simulators they should appreciate that at present, it is these endpoints that will most reliably reflect their performance. This, however, needs to be in the context of other current clinically agreed criteria for colonoscopy assessment end points, such as lesion detection rate and withdrawal time as well as therapeutic competency to be meaningful and relevant.

This review was limited by the degree of variation in methodology between included studies. This was particularly true for group assignment when examining construct validity. Participants were assigned to groups according to the number of previous colonoscopies performed. This raises two issues. First, it could be argued that the number of previous colonoscopies is not a good measure of experience. In effect, using this parameter may influence the reliability of any validation study. A more appropriate measure of experience could be provided by reviewing scores for previously completed Direct Observation of

Procedure or Skills (DOPS). Second, the definition of "novice," "intermediate," or "experience" operators was highly variable between studies (Table 7). For example, one study [28] considered participants with >200 previous colonoscopies as "experienced," whereas another [23] considered >1,000 to be an appropriate figure. This may not have adversely influenced construct validity within a study but made interstudy quantitative analysis unreliable.

Although it is tempting to be enthusiastic about new technology and its potential applications, this review shows us that validation of colonoscopic simulation is not yet complete. To fully validate VR models, more study of criterion validity is needed. This type of validation provides strong evidence but often is difficult to measure. Future work could compare VR scores with a previously validated assessment score, such as the DOPS. Further validation studies also may be required to evaluate each individual case/module within VR simulators using relevant endpoints. Standardized methodology between studies would provide stronger evidence. Preliminary results for animal composite colonoscopy simulation are encouraging for content, construct, and criterion validity. Given its ability to distinguish clearly between intermediate and experienced users, it may have the advantage of providing a useful training tool for senior endoscopic trainees. At present, trainees can use this type of simulation to support their clinical training rather than replace it. There is some way to go before these simulators should be incorporated into compulsory training. Further research on therapeutic colonoscopy simulation may be needed if these tools are to be used to assess progression into independent practice.

Disclosures Dr. Neil Warren, Dr. Neil Hawkes and Dr. Sunil Dolwani were part of the team that designed and constructed the WIMAT endoscopy suitcase. Mr J. Ansell, Mr. Jared Torkington, Mr. John Mason, Dr. Neil Hawkes, Prof. Peter Donnelly, and Dr. Sunil Dolwani have no conflicts of interest or financial ties to disclose. As lead author Mr J. Ansell was jointly sponsored by The Royal College of Surgeons of England Research Fellowship Grant and The Life Long Learning Grant.

References

1. Chesser S, Bowman K, Phillips H (2002) The European Working Time Directive and the training of surgeons. BMJ 325(7362):S69
2. Donohoe CL, Sayana MK, Kennedy MT, Niall DM (2010) European Working Time Directive: implications for surgical training. Ir Med J 103(2):57–59
3. Morris-Stiff GJ, Sarasin S, Edwards P, Lewis WG, Lewis MH (2005) The European Working Time Directive: one for all and all for one? Surgery 137(3):293–297
4. Waurick R, Weber T, Bröking K, Van Aken H (2007) The European Working Time Directive: effect on education and clinical care. Curr Opin Anaesthesiol 20(6):576–579
5. Jiang G, Chen H, Wang S, Zhou Q, Li X, Chen K, Sui X (2011) Learning curves and long-term outcome of simulation-based thoracentesis training for medical students. BMC Med Educ 11:39
6. Reynolds T, Kong ML (2010) Shifting the learning curve. BMJ. doi:10.1136/bmj.c6260. Dec 2, 2010
7. Stather DR, Maceachern P, Rimmer K, Hergott CA, Tremblay A (2011) Assessment and learning curve evaluation of endobronchial ultrasound skills following simulation and clinical training. Respirology 16(4):698–704
8. Koch AD, Buzink SN, Heemskerk J, Botden SM, Veenendaal R, Jakimowicz JJ, Schoon EJ (2008) Expert and construct validity of the Simbionix GI Mentor II endoscopy simulator for colonoscopy. Surg Endosc 22(1):158–162
9. The Joint Advisory Group on GI Endoscopy (JAG) (2011) Who we are. http://www.thejag.org.uk/aboutus/whoweare.aspx. Accessed 25 Aug 2011
10. Desilets DJ, Banerjee S, Barth BA, Kaul V, Kethu SR, Pedrosa MC, Pfau PR, Tokar JL, Varadarajulu S, Wang A, Wong Kee Song LM, Rodriguez SA, ASGE Technology Committee (2011) Endoscopic simulators. Gastrointest Endosc 73(5):861–867
11. Feldman LS, Sherman V, Fried GM (2004) Using simulators to assess laparoscopic competence: ready for widespread use? Surgery 135(1):28–42
12. American College of Surgeons Education Division (2011) Technical skills education surgery. http://www.facs.org/education/technicalskills/faqs/faqs.html. Accessed 25 Aug 2011
13. Sedlack RE (2011) Validation process for new endoscopy teaching tools. Tech Gastrointest Endosc 13:151–154
14. Meeker WQ, Escobar LU (1998) Statistical methods for reliability data. Wiley, Hoboken
15. Cortina JM (1993) What is coefficient alpha? An examination of theory and applications. J Appl Psychol 78(1):98–104
16. PRISMA (2011) PRISMA transparent reporting of systematic reviews and meta-analyses. http://www.prisma-statement.org/statement.htm. Accessed 25 Aug 2011
17. Google (2011) Google translate. http://translate.google.com/#. Accessed 25 Aug 2011
18. Centre for EBM Levels of Evidence (2011) Oxford Centre for Evidence-Based Medicine—levels of evidence. http://www.cebm.net/index.aspx?o=10252009. Accessed 25 Aug 2011
19. Fayez R, Feldman LS, Kaneva P, Fried GM (2010) Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters. Surg Endosc 24(5):1060–1065
20. Felsher JJ, Olesevich M, Farres H, Rosen M, Fanning A, Dunkin BJ, Marks JM (2005) Validation of a flexible endoscopy simulator. Am J Surg 189(4):497–500
21. Grantcharov TP, Carstensen L, Schulze S (2005) Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. JSLS 9(2):130–133
22. Haycock AV, Bassett P, Bladen J, Thomas-Gibson S (2009) Validation of the second-generation Olympus colonoscopy simulator for skills assessment. Endoscopy 41(11):952–958
23. Koch AD, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ (2008) A second-generation virtual reality simulator for colonoscopy: validation and initial experience. Endoscopy 40(9):735–738
24. Datta V, Mandalia M, Mackay S, Darzi A (2002) The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. Surg Endosc 16(10):1459–1463
25. MacDonald J, Ketchum J, Williams RG, Rogers LQ (2003) A lay person versus a trained endoscopist: Can the preop endoscopy simulator detect a difference? Surg Endosc 17(6):896–898
26. Mahmood T, Darzi A (2003) A study to validate the colonoscopy simulator. Surg Endosc 17(10):1583–1589
27. Sedlack RE, Kolars JC (2003) Validation of a computer-based colonoscopy simulator. Gastrointest Endosc 57(2):214–218

28. Moorthy K, Munz Y, Orchard TR, Gould S, Rockall T, Darzi A (2004) An innovative method for the assessment of skills in lower gastrointestinal endoscopy. Surg Endosc 18(11):1613–1619
29. Woo HS, Kim WS, Ahn W, Lee DY, Yi SY (2008) Haptic interface of the KAIST-Ewha colonoscopy simulator II. IEEE Trans Inf Technol Biomed 12(6):746–753
30. Sedlack RE, Baron TH, Downing SM, Schwartz AJ (2007) Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 102(1):64–74
31. Carter FJ, Schijven MP, Aggarwal R, Grantcharov T, Francis NK, Hanna GB, Jakimowicz JJ (2005) Work Group for Evaluation and Implementation of Simulators and Skills Training Programmes. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19(12):1523–1532
32. Satava RM, Gallagher AG, Pellegrini CA (2003) Surgical competence and surgical proficiency: definitions, taxonomy, and metrics. J Am Coll Surg 196:933–937
33. Williams CB (1998) Endoscopy teaching: time to get serious. Gastrointest Endosc 47(5):429–430
34. Hammoud MM, Nuthalapaty FS, Goepfert AR, Casey PM, Emmons S, Espey EL, Kaczmarczyk JM, Katz NT, Neutens JJ, Peskin EG, Association of Professors of Gynecology and Obstetrics Undergraduate Medical Education Committee (2008) To the point: medical education review the role of simulators in surgical training. Am J Obstet Gynaecol 199(4):338–343
35. Miskovic D, Wyles SM, Ni M, Darzi AW, Hanna GB (2010) Systematic review on mentoring and simulation in laparoscopic colorectal surgery. Ann Surg 252(6):943–951
36. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ (2010) A critical review of simulation based medical education. Med Educ 44:50–63
37. Gurusamy K, Aggarwal R, Palanivelu L, Davidson BR (2008) Systematic review of randomized controlled trials on the effectiveness of virtual reality training for laparoscopic surgery. Br J Surg 95(9):1088–1097
38. Fisher L, Ormonde DG, Riley RH, Laurence BH (2010) Endoscopic skills training in a simulated clinical setting. Simul Healthc 5(4):232–237
39. Haycock AV, East JE, Swain D, Thomas-Gibson S (2010) Development of a novel esophageal stricture simulation. Dig Dis Sci 55(2):321–327
40. Haycock AV, Youd P, Bassett P, Saunders BP, Tekkis P, Thomas-Gibson S (2009) Simulator training improves practical skills in therapeutic GI endoscopy: results from a randomized, blinded, controlled study. Gastrointest Endosc 70(5):835–845
41. Hassan I, Osei-Agymang T, Radu D, Gerdes B, Rothmund M, Fernández ED (2008) Simulation of laparoscopic surgery—four years' experience at the Department of Surgery of the University Hospital Marburg. Wien Klin Wochenschr 120(3–4):70–76
42. Maschuw K, Hassan I, Bartsch DK (2010) Surgical training using simulator Virtual Reality. Chirurg 81(1):19–24
43. Maagaard M, Sorensen JL, Oestergaard J, Dalsgaard T, Grantcharov TP, Ottesen BS, Larsen CR (2011) Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. Surg Endosc 25(3):722–727
44. Sturm LP, Windsor JA, Cosman PH, Cregan P, Hewett PJ, Maddern GJ (2008) A systematic review of skills transfer after surgical simulation training. Ann Surg 248(2):166–179
45. Ahlberg G, Hultcrantz R, Jaramillo E, Lindblom A, Arvidsson D (2005) Virtual reality colonoscopy simulation: a compulsory practice for the future colonoscopist? Endoscopy 37:1198–1204
46. Sedlack RE, Kolars JC, Alexander JA (2004) Computer simulation training enhances patient comfort during endoscopy. Clin Gastroenterol Hepatol 2:348–352
47. Sedlack RE, Kolars JC (2004) Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. Am J Gastroenterol 99:33–37
48. Cohen J, Cohen SA, Vora KC, Xue X, Burdick JS, Bank S, Bini EJ, Bodenheimer H, Cerulli M, Gerdes H, Greenwald D, Gress F, Grosman I, Hawes R, Mullin G, Schnoll-Sussman F, Starpoli A, Stevens P, Tenner S, Villanueva G (2000) Multicentre, randomised, controlled trial of virtual-reality simulator training in acquisition of competency in colonoscopy. Gastrointest Endosc 64:361–368
49. Aabakken L, Adamsen S, Kruse A (2000) Performance of a colonoscopy simulator: experience from a hands on course. Endoscopy 32:911–913
50. The Chamberlain Group (2011) General and gastrointestinal: lower GI. http://www.thecgroup.com/products/lower-gi. Accessed 25 Aug 2011
51. Simbionix (2011) GI mentor. http://simbionix.com/simulators/gi-bronch-gi-mentor/. Accessed 25 Aug 2011
52. WIMAT (2011) The WIMAT colonoscopy suitcase. http://www.cardiff.ac.uk/pgmde/sections/wimat/simulationtrainingsupport/simulation-training-support.html. Accessed 25 Aug 2011

Surg Endosc (2013) 27:1468–1477
DOI 10.1007/s00464-012-2631-7

REVIEWS

Is motion analysis a valid tool for assessing laparoscopic skill?

John D. Mason · James Ansell · Neil Warren · Jared Torkington

Received: 1 June 2012 / Accepted: 21 September 2012 / Published online: 12 December 2012

Abstract
Background The use of simulation for laparoscopic training has led to the development of objective tools for skills assessment. Motion analysis represents one area of focus. This study was designed to assess the evidence for the use of motion analysis as a valid tool for laparoscopic skills assessment.
Methods Embase, MEDLINE and PubMed were searched using the following domains: (1) motion analysis, (2) validation and (3) laparoscopy. Studies investigating motion analysis as a tool for assessment of laparoscopic skill in general surgery were included. Common endpoints in motion analysis metrics were compared between studies according to a modified form of the Oxford Centre for Evidence-Based Medicine levels of evidence and recommendation.
Results Thirteen studies were included from 2,039 initial papers. Twelve (92.3 %) reported the construct validity of motion analysis across a range of laparoscopic tasks. Of these 12, 5 (41.7 %) evaluated the ProMIS Augmented Reality Simulator, 3 (25 %) the Imperial College Surgical Assessment Device (ICSAD), 2 (16.7 %) the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD), 1 (8.33 %) the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT) and 1 (8.33 %) the Robotic and Video Motion Analysis Software (ROVIMAS). Face validity was reported by 1 (7.7 %) study each for ADEPT and ICSAD. Concurrent validity was reported by 1 (7.7 %) study each for ADEPT, ICSAD and ProMIS. There was no evidence for predictive validity.
Conclusions Evidence exists to validate motion analysis for use in laparoscopic skills assessment. Valid parameters are time taken, path length and number of hand movements. Future work should concentrate on the conversion of motion data into competency-based scores for trainee feedback.

Keywords Quality Control · Surgical < technical · Education

J. D. Mason
Cardiff University School of Medicine, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK
e-mail: masonjd85@gmail.com

J. Ansell (✉)
Royal College of Surgeons of England Research Fellow, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK
e-mail: ansellj@cf.ac.uk

N. Warren
Postgraduate Deanery Wales, Welsh Institute of Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, UK

J. Torkington
Department of Colorectal Surgery, University Hospital of Wales, Heath Park, Cardiff, UK

Subjective methods of trainee assessment are no longer adequate for surgical training [1]. Reduced working hours [2, 3], increased demands from the political sector [4] and financial pressures [5] mean that more objective measures are required. Surgical simulation is an effective tool for training and assessment. Simulators can reduce learning curves outside the operating theatre in a pressure-free environment, without requiring formal supervision [6]. Studies show that skills acquired during simulation training are transferable to the operating room [7]. Simulation in laparoscopic training refers to a wide range of devices from simple box trainers [8], cadaveric models [9], live animals

[10], to complex virtual reality (VR) systems (e.g. MIST-VR®, LapSim® ProMIS™, and LapMentor™) [11–14]. This has led to the development of simulator assessment tools which include motion analysis.

Motion analysis allows assessment of surgical dexterity using parameters that are extracted from movement of the hands or laparoscopic instruments [15]. Several different motion analysis systems have been developed (Table 1). This can be inbuilt within a simulator (e.g. ProMIS™) or as a separate device, enabling flexible use (e.g. Imperial College Surgical Assessment Device, ICSAD) [16]. Objective assessment of laparoscopic skill could be carried out using motion analysis if endpoints for each parameter are quantified according to pre-defined levels of experience. The conversion of motion analysis data into competency-based scores or indexes could provide a valuable source of trainee feedback [17]. This is an automatic and instant process [18]. Feedback could be useful on two levels, firstly by providing a quantitative index to define varying levels of experience, which trainees can work towards. Secondly, it could serve as evidence of professional development that is assessed at annual progress reviews. Before motion analysis can be used to assess laparoscopic competence, the technology and metrics measured must first be validated [19].

Validation of any new method for training or assessment is a critical step [20]. This is the extent to which an instrument measures what it was designed to measure [21, 22]. The process should begin by defining a "construct", which defines the underlying trait for which a new training tool is designed [20]. The more forms of validity (Table 2) that are demonstrated, the stronger the overall argument [20].

The aims of this systematic review are to provide an overview of the different motion analysis technologies available for the assessment of laparoscopic skill, and to identify the evidence for their validity.

Methods

Data resources and search criteria

A systematic review was performed according to the Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) guidelines [23]. The literature search was conducted using the following databases: Embase Classic + Embase (1947 to 2011 week 38), MEDLINE (1947 to present) and PubMed. For each database we searched three domains of exploded MeSH keyword terms. The general terms for each domain were (1) motion analysis, (2) validation and (3) laparoscopy. Where a keyword mapped to further subject headings, those considered relevant were also exploded to maximise coverage of the literature. Studies published in a foreign language were translated into English [24]. The last search date was 29 September 2011. This search strategy was undertaken by two independent reviewers, and articles retrieved according to the inclusion criteria. Articles arising from cross-referencing were also included. Duplicate articles and those

Table 1 Summary of motion analysis systems available for assessment of laparoscopic skill in general surgery

Motion analysis system	Description
Advanced Dundee Endoscopic Psychomotor Tester (ADEPT)	Consists of a dome enclosing a defined workspace that contains a target plate. Trainees are instructed to undertake up to four tasks using the target plate, including flicking a switch and turning a dial. Excessive contact with the plate or contact outside of the plate is measured as an error, recorded in seconds. Total time required to execute a task is also recorded [27].
Hiroshima University Endoscopic Surgical Assessment Device (HUESAD)	Consists of optical scale sensors, micro-encoders, an experimental table and monitor, which are connected to a computer. This enables the movement of the instrument tips to be tracked while a task is being performed. It is possible to measure two rotation angle parameters, one distance parameter and time taken [29].
Imperial College Surgical Assessment Device (ICSAD)	Utilises electromagnetic sensors placed on the dorsum of a trainee's hands, allowing hand movements to be tracked. Allows use within simulated and operating theatre environments. Data are produced by custom-built software [16].
ProMIS Augmented Reality Simulator	Trainees are able to use laparoscopic instruments to interact with the virtual environment, also including haptic feedback. Movements in space are tracked by a computer, which derives the performance metrics (time, smoothness and path length). Tasks include basic laparoscopic movements such as camera navigation as well as more complex tasks such as sharp dissection [13].
Robotic Video and Motion Analysis Software (ROVIMAS)	Translates three-dimensional coordinate data from the Isotrak II motion tracking device (Polhemus Inc., Colchester, VT) into useful motion parameters, e.g. time taken, path length and number of movements for each hand [18].

Table 2 Overview of validity types (adapted from Moorthy et al. [44])

Approach	Type of validity	Description	Method of examination
Subjective	Face validity	The extent to which the test or task resembles the real-life equivalent	Expert questionnaires
	Content validity	The extent to which the domain that is being measured is actually measured by the assessment tool	Expert questionnaires
Objective	Construct validity	The extent to which a test measures the trait that it purports to measure (for example the extent to which a test or task discriminates between various levels of expertise)	Measurement of relevant parameters between groups of variable experience
	Criterion validity		
	1. Concurrent validity	The extent to which the results of the assessment tool correlate with the gold standard for that particular domain	Comparison with patient-based data
	2. Predictive validity	The ability of the test or task to predict future performance	Correlation between test or task scores with future performance scores

clearly unrelated to the inclusion criteria were excluded. Any disagreements between the reviewers were referred to a third party.

Inclusion and exclusion criteria

All studies investigating motion analysis as a valid tool for assessment of laparoscopic skill in general surgery were included. Inclusion criteria included: sufficient detail of motion analysis technology used (including information regarding the precise motion metrics measured), description of the tasks being investigated and the type of validity measured. Studies that validated laparoscopic simulators for which motion analysis did not form the primary method of assessment were excluded. Furthermore, studies were excluded if they were validating assessment tools in specialities other than general surgery and/or if motion analysis was being validated for *laparoscopic training* rather than assessment. Evidence validating motion analysis for laparoscopic training is limited, and its inclusion would lead to further study design heterogeneity. Review articles and conference abstracts were also excluded.

Outcome measures and analysis

Each of the studies included was rated according to a modified form of the Oxford Centre for Evidence-Based Medicine (CEBM) levels of evidence and recommendation [25, 26]. Information was extracted from each study in accordance to the inclusion criteria. Common endpoints between studies were identified and compared when statistically significant results were reported, the principle summary statistic being the difference in means or medians. It was judged that the data were not suitable for meta-analysis due to study design heterogeneity.

Results

The primary search identified 2,039 records. Three hundred and eighty-eight duplicates were removed, and the remaining 1,651 abstract records screened for relevance. Following this process, 1,522 records were excluded and 129 full-text articles obtained. Full-text review excluded a further 124 studies, while cross-referencing identified 8 studies. At the end of this process, 13 studies were included for review (Fig. 1). These studies investigated four different motion analysis devices: the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT; two studies [27, 28]),

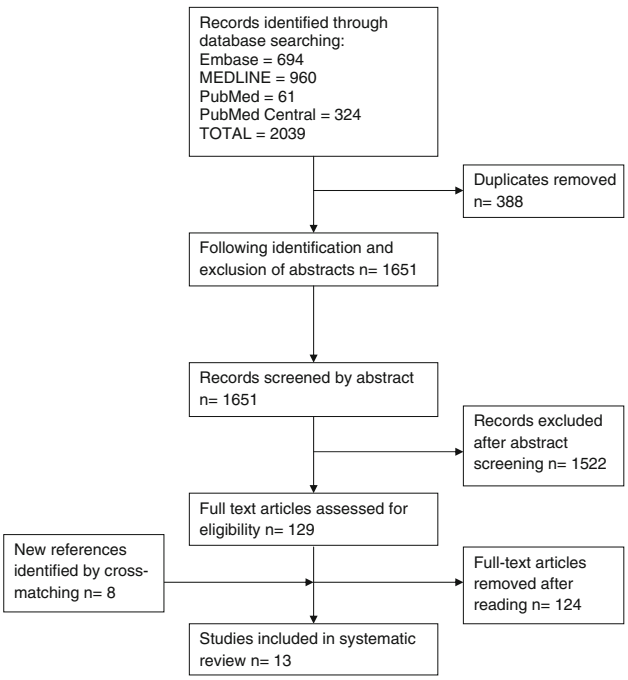


Fig. 1 PRISMA [23] flow diagram for selection of studies

Table 3 Studies included in review

Motion analysis	Authors
ADEPT	Macmillan et al. [27]
	Francis et al. [28]
HUESAD	Egi et al. [29]
	Tokunaga et al. [30]
ICSAD	Smith et al. [9]
	Moorthy et al. [31]
	Xeroulis et al. [32]
ProMIS	Van Sickle et al. [13]
	Broe et al. [33]
	Oostema et al. [34]
	Pellen et al. [35]
	Pellen et al. [36]
ROVIMAS	Aggarwal et al. [18]

the Hiroshima University Endoscopic Surgical Assessment Device (HUESAD; two studies [29, 30]), the Imperial College Surgical Assessment Device (ICSAD; three studies [9, 31, 32]), the ProMIS Augmented Reality Simulator (five studies [13, 33–36]) and the Robotic and Video Motion Analysis Software (ROVIMAS; one study [18]) (Table 3). No randomized controlled trials (RCTs) were identified. Twelve studies were graded at level 2b evidence [9, 13, 18, 28–36], and one study at level 3 [27].

Construct validity

Construct validity was examined in 12 (92.3 %) studies [9, 13, 18, 28–36]. There was a large degree of variation between studies, in terms of both group allocation and methodology (Table 4). Comparison between common endpoints (Table 5) was made in order to provide the following levels of recommendation (Table 6):

ADEPT: One study confirmed construct validity for the error score endpoint [28], when comparing novices and experts (level 3 recommendation).

HUESAD: Two studies established construct validity for the following endpoints: time taken to complete task [29, 30] (level 2 recommendation), deviation from ideal vertical and horizontal planes [29] and approaching time [30] (level 3 recommendation) when comparing novices and experts during a navigation task.

ICSAD: Three studies reported construct validity for the following endpoints: time (stage 1, 2 and 4 [9], tasks 1, 2, 3 and 4 [32]), number of hand movements (stage 1, 2 and 4 [9], task 1, 3 and 4 [32]) and path length (stage 1 and 2 [9], task 1 and 4 [32]) when comparing novices, intermediates and experts in the following tasks: laparoscopic cholecystectomy (LC) [9] and fundamentals of laparoscopic surgery (FLS) tasks [32] (all level 2 recommendations). Moorthy et al. [31] reported construct validity for time and path length in their laparoscopic suturing task when novices were compared with intermediates, and intermediates with experts. Two of the studies also demonstrated construct validity of overall expert rating scales that were used alongside motion analysis (level 2 recommendation) [31, 32].

ProMIS: Five studies established construct validity for the following endpoints: time [13, 33–36], path length [13, 34–36], smoothness of movement [13, 34–36] (level 2 recommendation) and number of hand movements [33] (level 3 recommendation) when comparing novices versus experts [13], novices versus intermediates versus experts [34, 35] or medical students/preregistration house officers (PRHOs) versus senior house officers versus surgical trainees versus consultants [33, 36] in various laparoscopic bench tasks. The tasks included suturing [13], orientation [33, 34, 36], object positioning [34, 36], knot tying [34] and sharp dissection [34–36].

ROVIMAS: One study confirmed construct validity for the following endpoints: time (overall, stage 1, 2 and 3), number of hand movements (stage 1) and path length (stage 1), when comparing novices and experts in a real-life LC [18] (level 3 recommendation). Number of hand movements and path length were both unable to distinguish between novices and experts in clipping and cutting the cystic duct (stage 2) and artery (stage 3), or during dissection (stage 4) [18].

Other validity types

Face validity was reported by one study for ADEPT [27] and one study for ICSAD [9] (no data provided). Three studies reported concurrent validity [9, 27, 35]. Macmillan et al. state that for ADEPT a high correlation was seen between the number of perfect runs and blinded clinical assessments (Spearman's rho 0.74) [27]. Concurrent validity was also confirmed by one study for ICSAD [31], and ProMIS [35], through the observation that motion analysis metrics correlated with expert and global rating scores (ICSAD: path length, Spearman's rho –0.78, $p = 0.000$; ProMIS: time and path length, Spearman's rho 0.88, $p < 0.05$) (all level 3 recommendations). None of the 13 studies included in this systematic review investigated content or predictive validity.

Discussion

This study presents the evidence for the use of motion analysis in laparoscopic skills assessment. A previous review by van Hove et al. [15] assessed a range of objective tools available to assess surgical skill, including

Table 4 Summary of methods

Motion analysis	Reference	*n*	Groups	Validation	Task	Endpoints
ADEPT	Macmillan et al. [27]	10	10 HSTs	Predictive and face	10 repeats of ADEPT tasks (Table 1)	Execution time, plate error score and probe error score
	Francis et al. [28]	40	*N* = 20 HSTs *E* = 20 consultants	Construct	ADEPT tasks (Table 1)	Execution time, plate error score, probe error score and overall performance
HUESAD	Egi et al. [29]	37	*N* = 25 medical students (no exp) *E* = 12 surgeons (>100 lap procedures)	Construct	Navigation	Time and deviation from the ideal course in the vertical and horizontal planes
	Tokunaga et al. [30]	36	*N* = 20 medical students *E* = 16 (>50 lap procedures)	Construct	Navigation	Total time, approaching time and intermediate time*
ICSAD	Smith et al. [9]	15	*N* = 5 (<10 human LCs) *I* = 7 (10–100 human LCs) *E* = 3 (>100 human LCs)	Construct and face	LC (porcine model) subdivided into four stages	Time, distance travelled, speed of movement and number of hand movements
	Moorthy et al. [31]	26	*N* = 13 (<10 LCs, no lap suturing) *I* = 7 (10–50 LCs, <50 lap sutures) *E* = 6 (>100 lap suturing procedures)	Construct and concurrent	Laparoscopic suturing (laboratory based)	Time and distance travelled
	Xeroulis et al. [32]	26	*N* = 13 (PGY 1–3) *I* = 7 (PGY 4–5) *E* = 6 staff surgeons	Construct	FLS education modules (four tasks)***	Time, distance travelled, number of hand movements and expert rating scores
ProMIS	Van Sickle et al. [13]	10	*N* = 5 (medical students) *E* = 5 (HSTs, significant lap exp)	Construct	Laparoscopic suturing	Time, path length, smoothness of movement‡, and error score
	Broe et al. [33]	20	Group 1 = 7 PRHOs Group 2 = 6 SHOs Group 3 = 1 JSpRs Group 4 = 3 HSTs Group 5 = 3 consultants	Construct	Laparoscopic orientation	Time, number of movements and OSATS global scoring
	Oostema et al. [34]	47	*N* = 24 medical students *I* = 19 PGY 1–5 *E* = 3 consultants	Construct	Laparoscopic orientation, object positioning, sharp dissection and knot tying	Time, path length and smoothness of movement
	Pellen et al. [35]	30	*N* = 10 medical students (no lap exp) *I* = 10 HSTs (<50 lap procedures) *E* = 10 consultants (>100 lap procedures)	Construct and concurrent	Sharp dissection (laboratory based)	Time, path length, smoothness and procedure-specific rating scale
	Pellen et al. [36]	160	Group 1 = 53 medical students Group 2 = 28 BSTs Group 3 = 61 HSTs Group 4 = 18 consultants (>100 lap procedures)	Construct	Laparoscopic orientation, object positioning and sharp dissection	Time, path length and smoothness

Table 4 continued

Motion analysis	Reference	*n*	Groups	Validation	Task	Endpoints
ROVIMAS	Aggarwal et al. [18]	19	*N* = 6 (<10 human LCs) *E* = 13 (>100 human LCs)	Construct	LC (on patient)	Time, distance travelled and number of hand movements

BST basic surgical trainee; *HST* higher surgical trainee; *JSpR* junior specialist registrar; *SSpR* senior specialist registrar; *N* novice; *I* intermediate; *E* expert; *LC* laparoscopic cholecystectomy; *PGY* postgraduate year; *PRHO* preregistration house officer; *SHO* senior house officer; *OSATS* Objective Structured Assessment of Technical Skills

* Approaching time = time taken to move between two points in HUESAD navigation tasks. Intermediate time = total time − approaching time

** Information regarding grade/experience not available

*** Fundamentals of laparoscopic surgery (FLS) is a CD-ROM-based education module for hands-on skills-based training [46]

‡ Smoothness is defined as the number of times an instrument changes velocity during completion of a task [34]

motion analysis. However, this did not provide information regarding the precise surgical skill assessed, nor did it provide subsequent levels of recommendation. The authors included studies validating the TrEndo Tracking System, which so far has only been studied in obstetrics and gynaecology trainees [37, 38]. These studies have produced promising results, and we recommend further studies investigating its application within general surgery. Carter et al. [26] published consensus guidelines concerning evidence rating and subsequent levels of recommendation for evaluation and implementation of simulators and skills training programmes [25, 26]. The authors produced an alternative system due to the absence of published validation studies that have rigorous experimental methodology [26]. Our review utilises this version of the CEBM system, and actual levels of recommendation for each tool have been provided for the first time (Table 6).

This review reports construct validity for a range of different motion analysis metrics across three different training environments (VR [13, 33–36], laboratory based [9, 28–32] and the operating theatre [18]). The most commonly validated metrics were time to complete a task, path length and number of hand movements. One ICSAD study attempted to establish construct validity for velocity during a simulated porcine LC model [9]. Velocity is a function of time and path length, both of which were also measured. However, while velocity was found to largely lack construct validity, this was not the case for time and path length. Smith et al. explain this by stating that each movement made by experienced surgeons is more efficient, meaning that, while the speed of movements is not significantly quicker, instead they are more goal directed so that tasks are completed in less time [9]. Despite only being a metric measured by the ProMIS simulator, smoothness of movements was also consistently shown to discriminate between different levels of experience [13, 33–36].

Aggarwal et al. [39] state the importance of breaking down training and assessment into basic, intermediate and advanced stages. It could be suggested that ADEPT and HUESAD could be used to assess basic training as they utilise simple orientation and movement skills in a non-anatomical environment. ICSAD, ProMIS and ROVIMAS could be used to assess intermediate competence. There are animal tissue models and virtual reality simulators that exist for a range of general surgery procedures that could be used in conjunction with these motion analysis technologies. This has already been demonstrated in a porcine model for LC [9], and adaptations to the devices may enable their use in endoscopy training. The flexibility of use offered by ICSAD and ROVIMAS means that advanced competency could be assessed. Construct validity during a real-life LC has already been demonstrated for ROVIMAS [18].

This systematic review also showed that very few forms of validity are being examined apart from construct. The more forms of validity that are demonstrated, the stronger the overall argument for the use of a particular technology [20]. While two studies report face validity [9, 27], no expert rating data were provided to support this. It may not be possible to face-validate motion analysis technology, as any attempt to do so would be assessing the realism of the laparoscopic set-up instead. While it is important to establish construct validity for each endpoint and in every procedure that motion analysis may eventually be used to assess, its practical use in real-life assessment is limited. Predictive validity represents a more useful modality to investigate, and it is unfortunate that there have been no studies undertaken to investigate this.

The main limitation of this review is the degree of methodological variation between included studies, which prevented meta-analysis. The largest degree of variation was seen for group allocation, which was largely based on career grades, although most studies used further inclusion criteria within each grade based on varying levels of laparoscopic experience. This limitation is explained by the fact that number of procedures performed is a

Table 5 Summary of statistical outcomes for construct validity studies

Motion analysis	Reference	Group/task	Time	Velocity	Path length*	No. of hand movements	Smoothness	↕ Deviation	↔ Deviation	Error score	Approaching time	Int. time	Global rating
ADEPT	Francis et al. [28]	*N* vs. *E*	0.420	–	–	–	–	–	–	**0.007**	–	–	0.400
HUESAD	Egi et al. [29]	*N* vs. *E*: task 1	**<0.0001**	–	–	–	–	**0.009**	**0.0004**	–	–	–	–
		N vs. *E*: task 2	**<0.0001**	–	–	–	–	**0.0002**	**<0.0001**	–	–	–	–
	Tokunaga et al. [30]	*N* vs. *E*: task 1	**<0.0001**	–	–	–	–	–	–	–	**<0.0001**	>0.050	–
		N vs. *E*: task 2	**<0.0001**	–	–	–	–	–	–	–	**<0.0001**	>0.050	–
ICSAD	Smith et al. [9]	*N* vs. *I* vs. E: Calot's triangle	**<0.000**	NS	**0.0120**	**<0.000**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: clipping duct/artery	**<0.000**	NS	**0.002**	**<0.000**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: cutting duct/artery	NS	NS	NS	NS	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: dissection	**<0.000**	**0.032**	NS	**0.049**	–	–	–	–	–	–	–
	Moorthy et al. [31]	*N* vs. *I* vs. *E*	**0.000**	–	**0.000**	–	–	–	–	–	–	–	**0.000**
		N vs. *I*	**0.060**	–	**0.010**	–	–	–	–	–	–	–	–
		I vs. *E*	**0.001**	–	**0.000**	–	–	–	–	–	–	–	–
	Xeroulis et al. [32]	*N* vs. *I* vs. *E*: overall	–	–	–	–	–	–	–	–	–	–	**0.001**
		N vs. *I* vs. *E*: task 1	**0.001**	–	**0.025**	**0.050**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 2	**0.002**	–	0.286	0.297	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 3	**0.005**	–	0.121	**0.011**	–	–	–	–	–	–	–
		N vs. *I* vs. *E*: task 4	**0.001**	–	**0.001**	**0.002**	–	–	–	–	–	–	–
ProMIS	Van Sickle et al. [13]	*N* vs. *E*	**<0.0001**	–	**<0.0001**	–	**<0.0001**	–	–	–	–	–	
	Broe et al. [33]	1 vs. 4 and 5	**<0.050**	–	–	**<0.050**	–	–	–	–	–	–	–
	Oostema et al. [34]	*N* vs. *I* vs. *E*: camera navigation	**0.09**	–	0.900	–	**0.030**	–	–	–	–	–	–
		N vs. *I* vs. *E*: object positioning	**<0.010**	–	**0.010**	–	**<0.010**	–	–	–	–	–	–
		N vs. *I* vs. *E*: sharp dissection	**<0.010**	–	**<0.010**	–	**<0.010**	–	–	–	–	–	–
		N vs. *I* vs. *E*: knot tying	**<0.010**	–	**<0.010**	–	**<0.010**	–	–	–	–	–	–
	Pellen et al. [35]	*N* vs. *I* vs. *E*	**0.010**	–	**0.010**	–	**0.010**	–	–	0.143	–	–	**<0.050**
	Pellen et al. [36]	1 vs. 2 vs. 3 vs. 4 vs. 5: orientation	0.058	–	0.097	–	**0.008**	–	–	–	–	–	–
		1 vs. 2 vs. 3 vs. 4 vs. 5: object positioning	**0.000**	–	**0.000**	–	**0.000**	–	–	–	–	–	–
		1 vs. 2 vs. 3 vs. 4 vs. 5: sharp dissection	**0.000**	–	**0.000**	–	**0.000**	–	–	**0.001**	–	–	–
ROVIMAS	Aggarwal et al. [18]	*N* vs. *E*: overall	**0.036**	–	0.625	0.389	–	–	–	–	–	–	–
		N vs. *E*: Calot's triangle	**0.002**	–	**0.048**	**0.007**	–	–	–	–	–	–	–
		N vs. *E*: clip and cut duct	**0.013**	–	0.063	0.553	–	–	–	–	–	–	–
		N vs. *E*: clip and cut artery	**0.002**	–	0.119	0.204	–	–	–	–	–	–	–
		N vs. *E*: dissection	0.471	–	0.377	0.396	–	–	–	–	–	–	–

NS no significant difference (where *p* values not provided); *N* novice; *I* intermediate; *E* expert

– Denotes parameter not measured in study/data not provided

Bold indicates *p* value significant at 0.05 level

* Equivalent to distance moved by instrument

Table 6 Level of evidence and recommendation for each motion analysis device

Motion analysis device	Type of validity (level of evidence)			Recommendation level
	Face	Concurrent	Construct	
ADEPT	Yes (3) [27][§]	Yes (3) [27]	Yes, for error score (2b) [28]	Level 4 for face and concurrent validity Level 3 for construct validity endpoints
HUESAD	No	No	Yes, for time taken to complete task (2b) [29, 30] (*)	Level 2 for (*) construct validity endpoints; level 3 for (**) construct validity endpoints.
			Yes, for deviation from ideal vertical plane (2b) [29]	
			Yes, for deviation from ideal horizontal plane (2b) [29] (**)	
			Yes, for approaching time (2b) [30]	
ICSAD	Yes (2b) [9][§]	Yes (2b) [31]	Yes, for time taken to complete task (2b) [9, 31, 32]	Level 3 for face validity
			Yes, for number of hand movements (2b) [9, 31, 32]	All level 2 for construct validity endpoints
			Yes, for path length (2b) [9, 31, 32]	
ProMIS	No	Yes (2b) [35]	Yes, for time taken to complete task (2b) [13, 33–36]	Level 3 for concurrent validity
			Yes, for path length (2b) [13, 34–36] (*)	
			Yes, for smoothness (2b) [13, 34–36]	Level 2 for (*) construct validity endpoints; level 3 for (**) construct validity endpoints
			Yes, for number of hand movements (2b) [33] (**)	
ROVIMAS	No	No	Yes, for time taken to complete task (2b) [18]	All level 3 for construct validity endpoints
			Yes, for number of hand movements (2b) [18]	
			Yes, for path length (2b) [18]	

[§] No expert rating data provided; Macmillan et al. [27] state face validity is assured as the study utilised the same equipment used in minimal access surgery. Smith et al. [9] state face validity is assured due to the observation that there was little change in performance amongst members of the same group

non-objective measure of experience. A more objective approach to group allocation could have been made on the basis of Objective Structured Assessment of Technical Skills (OSATS) scoring. A further limitation is that the majority of the studies included compared groups across wide ranges of experience (e.g. novice versus intermediate versus expert), where outcomes may be largely dependent on the novice versus expert element of this analysis. Motion analysis must demonstrate the sensitivity to discriminate between all individual grades if it is to be used to assess laparoscopic competence.

Motion analysis does carry some limitations which require discussion. Firstly, many of the devices require calibration to account for individual physiological tremor. This may require technical support during each procedure. Additionally, there is the issue of cost, which may prevent widespread use across all training centres.

In order for motion analysis to be used as an assessment tool it must be shown to work in a real-life environment. While the feasibility of using motion analysis in a real-life operating theatre has been demonstrated for ICSAD [40] and ROVIMAS [18], the correlation between motion analysis assessment in the laboratory and its subsequent use within the operating theatre needs to be evaluated. Quantitative assessment outcomes must be shown to be equivalent between different training environments, otherwise the application of motion analysis to provide trainee feedback is undermined.

Using motion analysis in isolation may remove the user from the context of the operating theatre. As surgical competence is multimodal, it is important that assessment is not only based on specific outcomes (such as dexterity) but also global outcomes, such as task accuracy and outcome. This is made possible through the dual application of motion analysis alongside global checklists [e.g. Global Operative Assessment of Laparoscopic Skills (GOALS) and Objective Structured Assessment of Technical Skills (OSATS)] [41]. Furthermore, procedure-specific rating scales have also been developed to assess specific technical aspects of different operations, including LC [42] and Nissen fundoplication [43]. Using these systems, assessment can either occur "live", whilst a trainee is

undertaking a specific task [44]. Several studies included in this review included global rating scores, which were found to correlate with motion analysis metrics [18, 31, 35].

It has been suggested that surgery is 75 % decision-making and 25 % dexterity [45]. While motion analysis may provide a promising tool to assess dexterity, it cannot provide information on the numerous attributes that contribute to the other three-quarters of a good surgeon's skill set. Further work is needed to correlate motion analysis against similarly validated measurements of surgical decision-making in different scenarios.

Conclusions

We have demonstrated that there is evidence validating the use of motion analysis to assess laparoscopic skill. The most valid metrics appear to be time, path length and number of hand movements. More work is needed to establish predictive validity for each of these metrics. Future work should concentrate on the conversion of motion analysis data into competency-based scores or indices for trainee feedback.

Acknowledgments J. Ansell is currently funded by a Royal College of Surgeons England research fellowship grant.

Disclosures Authors John D. Mason, James Ansell, Neil Warren and Jared Torkington have no conflicts of interest or financial ties to disclose.

References

1. Darzi A, Smith S, Taffinder N (1999) Assessing operative skill. Using these systems, assessment. BMJ 318:887–888
2. Chesser S, Bowman K, Phillips H (2002) The European Working Time Directive and the training of surgeons. BMJ 325:S69
3. Donohoe C, Sayana M, Kennedy M, Niall D (2010) European Time Directive: implications for surgical training. Ir Med J 103:57–59
4. Smith R (1998) All changed, changed utterly: British medicine will be transformed by the Bristol case. BMJ 316:1917–1918
5. Bridges M, Diamond D (1999) The financial impact of teaching surgical residents in the operating room. Am J Surg 177:28–32
6. Munz Y, Kumar B, Moorthy K, Bann S, Darzi A (2004) Laparoscopic virtual reality and box trainers. Is one superior to the other? Surg Endosc 18:485–494
7. Sturm L, Windsor J, Cosman P, Cregan P, Hewett P, Maddern G (2008) A systematic review of skills transfer after surgical simulation training. Ann Surg 248(2):166–179
8. Chung J, Sackier J (1998) A method of objectively evaluating improvements in laparoscopic skills. Surg Endosc 12:1117–1120
9. Smith S, Torkington J, Brown T, Taffinder N, Darzi A (2002) Motion analysis. A tool for assessing laparoscopic dexterity in the performance of a laboratory-based laparoscopic cholecystectomy. Surg Endosc 16:640–645
10. Clayden G (1994) Teaching laparoscopic surgery. Preliminary training on animals is essential. BMJ 309:342
11. Wilson M, Middlebrooke A, Sutton C, Stone R, McCloy R (1997) MIST-VR: a virtual reality trainer for laparoscopic surgery assesses performance. Ann R Coll Surg Engl 79:403–404
12. Larson A (2001) An open and flexible framework for computer aided surgical training. Stud Health Technol Inform 81:263–265
13. Van Sickle K, McClusky DI, Gallagher A, Smith C (2005) Construct validation of the ProMIS simulator using a novel laparoscopic suturing task. Surg Endosc 19:1227–1231
14. McDougall E, Corica F, Boker J, Sala L, Stoliar G, Borin J, Chu F, Clayman R (2006) Construct validity testing of a laparoscopic surgical simulator. J Am Coll Surg 202(5):779–787
15. van Hove P, Tuijthof G, Verdaasdonk E, Stassen LP, Dankelman J (2010) Objective assessment of technical surgical skills. Br J Surg 97:972–987
16. Datta V, Mackay S, Darzi A, Gillies D (2001) Motion analysis in the assessment of surgical skill. Comput Methods Med Biomed Eng 4:515–523
17. Aggarwal R, Hance J, Darzi A (2004) Surgical education and training in the new millenium. Surg Endosc 18:1409–1410
18. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Papasavas P, Dosis A, Bello F, Darzi A (2007) An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg 245(6):992–999
19. Thijssen A, Schijven M (2010) Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? Am J Surg 199:529–541
20. Sedlack R (2011) Validation process for new endoscopy teaching tools. Tech Gastro Endosc 13:151–154
21. Feldman L, Sherman V, Fried G (2004) Using simulators to assess laparoscopic competence: ready for widespread use? Surgery 135:28–42
22. Technical skills education surgery (2011). http://www.facs.org/education/technicalskills/faqs/faqs.html. Accessed 8 Oct 2011
23. The PRISMA (2011). http://www.prisma-statement.org/index.htm. Accessed 1 Oct 2011
24. Google Translate (2011). http://translate.google.com/. Accessed 30 Sept 2011
25. Oxford Centre for Evidence Based Medicine—Levels of Evidence (2011). http://www.cebm.net/index.aspx?o=1025. Accessed 30 Sept 2011
26. Carter F, Schijven M, Aggarwal R, Grantcharov T, Francis N, Hanna G, Jakimowicz J (2005) Work group for evaluation and implementation of simulators and skills training programmes. Consensus guidelines for validation of virtual reality surgical simulators. Surg Endosc 19:1523–1532
27. Macmillan A, Cuschieri A (1999) Assessment of innate ability and skills for endoscopic manipulations by the advanced Dundee endoscopic psychomotor tester: predictive and concurrent validity. Am J Surg 177:274–277
28. Francis N, Hanna G, Cuschieri A (2002) The performance of master surgeons on the advanced Dundee endoscopic psychomotor tester. Arch Surg 137:841–844
29. Egi H, Okajima M, Yoshimitsu M, Ikeda S, Miyata Y, Masugami H, Kawahara T, Kurita Y, Kaneko M, Asahara T (2008) Objective assessment of endoscopic surgical skills by analyzing direction-dependent dexterity using the Hiroshima University endoscopic surgical assessment device (HUESAD). Surg Today 38:705–710
30. Tokunaga M, Egi H, Minoru H, Yoshimitsu M, Sumitani D, Kawahara T, Okajima M, Ohdan H (2011) Approaching time is important for assessment of endoscopic surgical skills. Min Invasive Ther Allied Technol 21:142–149
31. Moorthy K, Munz Y, Doris A, Bello F, Chang A, Darzi A (2004) Bimodal assessment of laparoscopic suturing skills: construct and concurrent validity. Surg Endosc 18:1608–1612
32. Xeroulis G, Dubrowski A, Leslie K (2009) Simulation in laparoscopic surgery: a concurrent validity study for FLS. Surg Endosc 23:161–165

33. Broe D, Ridgway P, Johnson S, Tierney K, Conlon K (2006) Construct validation of a novel hybrid surgical simulator. Surg Endosc 20:900–904
34. Oostema J, Abdel M, Gould J (2008) Time-efficient laparoscopic skills assessment using an augmented-reality simulator. Surg Endosc 22:2621–2624
35. Pellen M, Horgan L, Barton J, Attwood S (2009) Laparoscopic surgical skills assessment: can simulators replace experts? World J Surg 33:440–447
36. Pellen M, Horgan L, Barton J, Attwood S (2009) Construct validity of the ProMIS laparoscopic simulator. Surg Endosc 23:130–139
37. Chmarra M, Kolkman W, Jansen F, Grimbergen C, Dankelman J (2007) The influence of experience and camera holding on laparoscopic instrument movements measured with the TrEndo tracking system. Surg Endosc 21:2069–2075
38. Chmarra M, Klein S, de Winter J, Jansen F-W, Dankelman J (2010) Objective classification of residents based on their psychomotor laparoscopic skills. Surg Endosc 24:1031–1039
39. Aggarwal R, Moorthy K, Darzi A (2004) Laparoscopic skills training and assessment. Br J Surg 91:1549–1558
40. Doris A, Aggarwal R, Bello F, Moorthy K, Munz Y, Gillies D, Darzi A (2005) Synchronized video and motion analysis for the assessment of procedures in the operating theater. Arch Surg 140:293–299
41. Martin J, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchinson C, Brown M (1997) Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. Br J Surg 84:273–278
42. Eubanks T, Clements R, Pohl D, Williams N, Schaad D, Horgan S, Pellegrini C (1999) An objective scoring system for laparoscopic cholecystectomy. J Am Coll Surg 189:566–574
43. Dath D, Regehr G, Birch D, Schlachta C, Poulin E, Mamazza J, Reznick R, MacRae H (2004) Toward reliable operative assessment: the reliability and feasibility of videotaped assessment of laparoscopic technical skills. Surg Endosc 18(12):1800–1804
44. Moorthy K, Munz Y, Sarker S, Darzi A (2003) Objective assessment of technical skills in surgery. BMJ 327:1032–1037
45. Spencer F (1978) Teaching and measuring surgical techniques—the technical evaluation of competence. Bull Am Coll Surg 63:9–12
46. FLS program (2011). http://www.flsprogram.org. Accessed 9 Oct 2011

Original article

doi:10.1111/j.1463-1318.2012.03122.x

The WIMAT colonoscopy suitcase model: a novel porcine polypectomy trainer

J. Ansell*, K. Arnaoutakis*, S. Goddard*, N. Hawkes†, R. Leicester‡, S. Dolwani§, J. Torkington¶ and N. Warren*

*Welsh Institute for Minimal Access Therapy (WIMAT), Heath Park, Cardiff CF14 4UJ, UK, †Royal Glamorgan Hospital, Llantrisant, Cardiff CF72 8XR, UK, ‡St George's Hospital (London), Blackshaw Road, London SW17 0QT, UK, §University Hospital Llandough, Penlan Road, Penarth, Cardiff CF64 2XX, UK and ¶University Hospital of Wales, Heath Park, Cardiff CF14 4XN, UK

Received 16 December 2011; accepted 23 February 2012; Accepted Article online 2 June 2012

Abstract

Aim Simulation allows the acquisition of complex skills within a safe environment. Endoscopic polypectomy has a long learning curve. Our novel polypectomy simulator may be a useful adjunct for training. The aim of this study was to assess its content validity.

Method The Welsh Institute for Minimal Access Therapy (WIMAT) endoscopy suitcase was designed to simulate colonic polypectomy. Participants from regional and national courses were recruited into the study. Each undertook a standardized simulated polypectomy and completed a seven-point Likert scale questionnaire examining its realism.

Results In all, 17 participants completed the questionnaire: 15 (88.2%) gastroenterologists, one (5.9%) colorectal surgeon and one (5.9%) experienced endoscopic nurse specialist. Of the gastroenterologists, seven (46.7%) were consultants and eight (53.3%) were senior trainees or Post CCT (Certificate of Completion of Training) fellows. The mean number of real-life polypectomies performed by the cohort was 156 (95% CI 35–355). The highest scores were for 'mucosal realism' (median score 6.0, $P = 0.001$), 'endoscopic snare control' (median score 6.0, $P = 0.001$), 'handling the polyp' (median score 6.0, $P = 0.001$) and 'raising mucosa' (median score 6.0, $P < 0.001$). Of the 15 parameters examined only three were not statistically significant in favour of the simulator. These were 'anatomical realism of sessile polyps', 'resistance of scope movement' and 'paradoxical motion'. The overall score for the simulation was 6.0 ($P < 0.001$). There was no significant difference between the level of difficulty of the simulator compared with real life (median score 4.0, $P = 0.559$).

Conclusion The WIMAT colonoscopy suitcase model has excellent content validity for several parameters. This may have potential applications in medical training and assessment.

Keywords Simulation, polypectomy, training, *ex vivo*, colonoscopy

What is new in this paper?

This paper details a new way of simulating endoscopic polypectomy for use in medical training. This is a novel *ex vivo* animal model with the potential to teach a complex procedure. The paper highlights its content validity.

Introduction

Simulation is widely used in medical training and assessment. The advantage of simulation is that it enables the development of practical skills in a controlled environment. This has the potential to improve patient care via a reduction in procedural complication rates.

The first endoscopic simulators were described between 1969 and 1970 [1,2]. Since then, there have been significant developments in the field. Simulators now exist with the capacity to teach several endoscopic procedures [3], including upper and lower gastrointestinal endoscopy, endoscopic retrograde cholangiopancreatography and endoscopic ultrasound. There are also different types of simulators available, ranging from bench models and *ex vivo* animal platforms to virtual

Correspondence to: Mr James Ansell, BSc, MBBCh, MRCS, Royal College of Surgeons Clinical Research Fellow in Simulation, Welsh Institute for Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff CF14 4UJ, UK.
E-mail: ansellj@cf.ac.uk

reality (VR) trainers [4–7]. For colonoscopy in particular there is a large weight of evidence which focuses on VR simulation [8–16]. Several VR validation studies have been conducted on commercially available products [8–16]. VR simulators [GI mentor II (Simbionix), Accutouch HT immersion, Olympus Endo Ts-1 2nd Generation] allow participants to practice multiple computer-based modules with varying degrees of difficulty. These simulators emphasize scope navigation and loop management with some limited capacity for therapeutic interventions. There is evidence to suggest that VR has the potential to oversimplify complex tasks [4,17–19]. This is illustrated by reports that these simulators are often unable to distinguish between participants with different levels of expertise [17–19]. Therefore, at present their most effective use could be for junior trainees with minimal experience. Others have questioned the usefulness of the multiple parameters that the VR simulators measure and their relevance to evaluating performance [4].

An alternative approach to VR is the use of *ex vivo* animal tissue simulation. Sedlack *et al.* [4] describe the validation of a novel bovine colonoscopy simulation for use in skills assessment and as an adjunct to senior training. The paper focuses on the skill of performing a colonoscopy but does not consider the simulation of therapeutic measures. There is a lack of current evidence for the validation of *ex vivo* animal models that focus on the area of therapeutic colonoscopic intervention.

Simulators should undergo a formal validation process before widespread adoption in training or assessment [20]. This allows inferences to be made regarding effectiveness and in justifying its investment [20]. The first stage of any validation process is to establish a construct [20]. This defines exactly what needs to be examined by a new training tool [20]. In colonoscopy, for example, the construct should clarify whether the simulator is being validated to assess or to teach its users [20]. The next step is to establish face and content validity. This is usually assessed by surveying experts regarding a given simulator's realism. Following this, the simulator should be assessed on its ability to distinguish between levels of expertise. This provides evidence for its construct validity. The final element to prove is criterion validity. This includes predictive validity (the ability of a tool to predict future performance) and concurrent validity (the correlation between the assessment tool and the 'gold standard') [21].

The aim of this study was develop an *ex vivo* animal model to simulate the procedure of colonic polypectomy. The Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase model is a novel porcine simulator which allows the participant to practice snare polypectomy of sessile and pedunculated (bleeding and non-bleeding) polyps. We report the evidence for its face and content validity.

Method

Porcine tissue

The WIMAT colonoscopy suitcase was developed at the WIMAT centre. Frozen porcine colonic specimens were sourced from a regional company (Fresh Tissue Supplies Ltd, Heathfield, East Sussex, UK) [22]. These originated from low risk category three animal by-products [23]. All animal samples were handled and disposed of according to a strict internal protocol. The specimens were initially defrosted and everted (Fig. 1a) to expose the mucosa, and the internal aspect of the colon underwent a standardized cleaning process. Three types of polyps were then constructed: sessile, pedunculated non-bleeding and pedunculated bleeding polyps. Sessile polyps were created by injecting a standardized volume of a polyp mix into the bowel submucosal layer. The polyp mix is a solution that solidifies at room temperature and does not break down when refrozen. The mix was warmed using a T.ARE heating magnetic stirrer (VELP®) to a temperature of 90°C [24]. Once liquefied, it was injected and cooled *in situ* to seal its position under the mucosa as a sessile polyp. The pedunculated non-bleeding polyps were

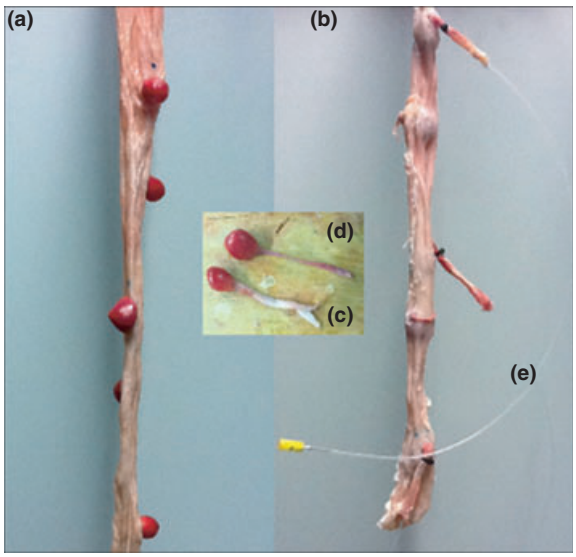


Figure 1 Porcine tissue: a, bowel everted (mucosa on outside, with pedunculated polyps attached); b, bowel inverted (mucosa on inside); c, pedunculated (bleeding) polyp; d, pedunculated (non-bleeding) polyp; e, catheter inserted into pedunculated polyp to simulate bleeding.

constructed by using a thin layer of sausage skin which was filled with a standardized volume of liquid polyp mix (Fig. 1d) and allowed to solidify. To create the pedunculated bleeding polyps this process was repeated and, in addition, a standardized length of porcine ureter was attached inside each polyp and cannulated with a small plastic catheter (Fig. 1c and e). This catheter was subsequently attached to a 50 ml Luer-Lock syringe containing simulated blood. All pedunculated polyps were attached to the bowel mucosa in a standardized way producing the view illustrated by Fig. 1a. The whole specimen was then inverted so that the polyps were transferred to the internal aspect of the lumen of the bowel (Fig. 1b).

The simulator casing

The porcine bowel was housed in a portable polymer suitcase (Storm Case Im2600, Hardigg®) with a hole made in one end to simulate the anus (Fig. 2a) [25]. This hole was cannulated with a 15 mm Ethicon XL

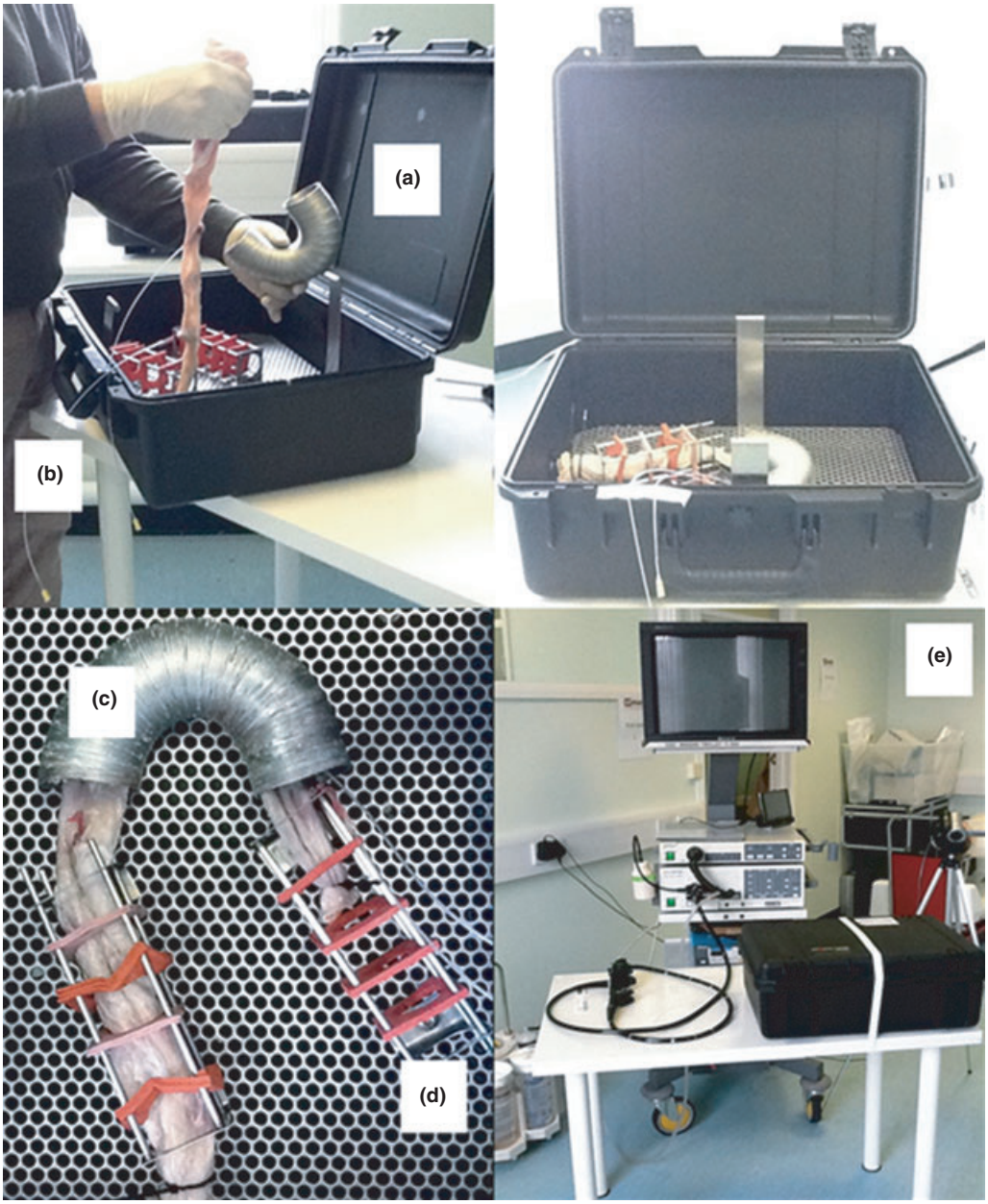


Figure 2 Set-up of the WIMAT endoscopy suitcase: a, plastic casing; b, catheter inserted into pedunculated polyp to simulate bleeding; c, exhaust piping to simulate sigmoid bend; d, foam segmentors were mounted onto metal rails; e, overview.

port (Johnson and Johnson) and secured internally with a plastic ring clip. This allowed the passage of the colonoscope into the suitcase and provided an airtight seal to enable insufflation of the colon. Inside the suitcase we placed a removable metal mesh base. This accommodated a crocodile clip which was connected to a diathermy unit, for use during the simulated polypectomy. Foam segmentors were mounted onto metal rails and the whole device was secured with wing nuts onto the mesh base (Fig. 2). The porcine bowel was then passed through the inside of the foam segmentors and a curved piece of 55 mm standard exhaust piping and then through a second set of segmentors (Fig. 2b) (this represented the sigmoid bend). The anal end of the specimen was attached to the Ethicon XL port using cable ties and the oral end was secured airtight. When the bowel was inflated, the foam segmentors indented on the serosal aspect of the bowel (to represent haustral folds) to give it a realistic luminal appearance (Fig. 3).

Endoscopic equipment

A standard Olympus Cf-Q140L colonoscope, CLV-U40 light, CV-240 processor and OEV 203 monitor were used for each procedure (Fig. 2e). All endoscopy equipment was dedicated to animal endoscopy teaching alone. Standardized endoscopic 25 mm snares (AcuSnare®) were supplied by Cook® (Wilson Cook Medical GI Endoscopy) [26]. The energy source was from Valley Lab^{TM} EZc and was placed on a cutting setting of 40 W [27]. If a mucosal lifting agent was required we used a water based dyed substance which was developed at the training centre. This contrasted with the colour of the polyps.

Validity testing

Participants were recruited from regional and national endoscopic training courses where the simulator was being demonstrated. All participants were experienced in the skill of colonic polypectomy. Each completed a snare polypectomy on a simulated pedunculated (bleeding/non-bleeding) and sessile polyp. All simulators and endoscopic equipment were standardized throughout. Following the procedure, each participant was asked to complete a 15-question realism survey based on a seven-point Likert scale. The questionnaire was adapted from that of Sedlack *et al.* [4] and reconstructed according to expert opinion at our research centre. Questions 1–13 of the survey were divided into the following three areas: visual realism, anatomical realism, mechanical realism (1 = strongly disagree, 4 = neutral, 7 = strongly agree). Question 14 focused on the overall degree of similarity between the simulated polypectomy and 'real-life' polypectomy (1 = strongly disagree, 4 = neutral, 7 = strongly agree). Question 15 compared the technical difficulty of human polypectomy with the simulation (1 = much easier, 4 = same, 7 = much more difficult).

Data analysis and power calculations

Assuming that the seven-point scales had a standard deviation of 1.0, 17 participants would give > 90% power to detect a difference of 1 point or more on the survey scale against a hypothetical mean of 4. The realism surveys were analysed using the Wilcoxon signed-rank test on a PASW Statistics 18 (SPSS) for non-parametric data. Median values from the seven-point scale were

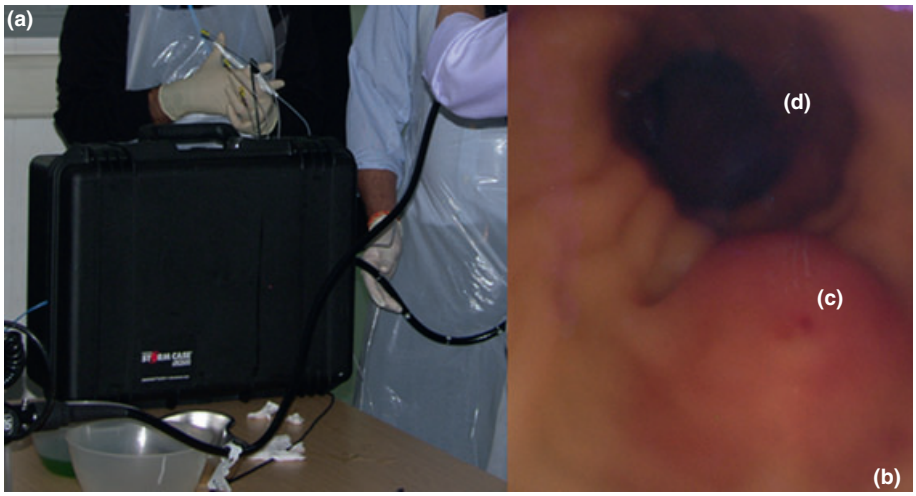


Figure 3 Simulator in use: a, external view; b, luminal view; c, sessile polyp; d, lumen.

compared with a hypothetical mean of 4 to determine statistical significance.

Results

General

A total of 17 participants (male: female ratio 14:3) completed the questionnaire: 15 (88.2%) gastroenterologists, 1 (5.9%) colorectal surgeon and one (5.9%) experienced endoscopic nurse specialist. Of the gastroenterologists, 7 (46.7%) were consultants, 8 (53.3%) were ST6-7 level or Post CCT fellows (Table 1). All participants were experienced in performing colonoscopy, polypectomy and polyp biopsies. The mean numbers of previous procedures performed by the cohort were 371 (95% CI 179–689) colonoscopies, 156 (95% CI 35–355) polypectomies and 165 (95% CI 42–360) biopsies (Table 1). The majority of the cohort had previous experience of using several different polypectomy simulators (Table 1).

Table 1 Experience levels of participants.

Level of participant	Number of participants (%)
Consultant	8 (47.1)
Post CCT	2 (11.8)
Senior trainee	6 (35.2)
Nurse specialist	1 (5.9)
Type of procedure	Number of real life procedures performed (mean with 95% CI)
Colonoscopies	371 (179–689)
Polypectomies	156 (35–355)
Biopsies	165 (42–360)
Type of simulation	Average previous number of times simulators used by delegates (mean with 95% CI)
Virtual reality	64 (3–180)
Animal model	4 (1–9)
Bench model	8 (1–19)

Realism survey

The highest scores were for 'mucosal realism' (median score 6.0, $P = 0.001$), 'endoscopic snare control' (median score 6.0, $P = 0.001$), 'handling the polyp' (median score 6, $P = 0.001$) and 'raising mucosa' (median score 6.0, $P < 0.001$) (Table 2). Six parameters scored a median score of 5 with statistically significant results (Table 2). These were 'endoscopic view' ($P = 0.001$), 'polyp realism' ($P < 0.001$), 'bleeding realism' ($P = 0.013$), 'haustral folds' ($P = 0.029$), 'anatomical realism of pedunculated polyps' ($P = 0.01$) and 'diathermy of the polyp' ($P = 0.026$) (Table 2). Of the 15 parameters examined only three were not statistically significant in favour of the simulator. These were 'anatomical realism of sessile polyps' ($P = 0.08$), 'resistance of scope movement' ($P = 0.406$) and 'paradoxical motion' ($P = 0.055$).

Table 2 Results of realism survey [median scores, 25%–75% interquartile range (IQR) for realism parameters using a seven-point Likert scale; Wilcoxon sign-rank testing to compare actual median with a hypothetical median neutral score of 4, $P < 0.05$].

Realism type	Realism aspect	Median score (IQR) ($n = 17$)	P
Visual	Mucosal realism	6.0 (5.0–6.0)	0.001
	Endoscopic view	5.0 (5.0–6.0)	0.001
	Polyp realism	5.0 (5.0–6.0)	<0.001
	Bleeding realism	5.0 (4.0–6.0)	0.013
	Haustral folds	5.0 (4.0–5.0)	0.029
Anatomical	Pedunculated polyps	5.0 (4.0–6.0)	0.010
	Sessile polyps	4.0 (4.0–5.0)	0.088
	Resistance to scope	4.0 (4.0–5.0)	0.406
	Paradoxical motion	4.0 (3.0–6.0)	0.055
	Snare control	6.0 (5.0–6.0)	0.001
Mechanical	Handling the polyp	6.0 (5.0–6.0)	0.001
	Diathermy of polyp	5.0 (4.0–6.0)	0.026
	Raising mucosa	6.0 (5.0–6.0)	<0.001
	Overall simulation	6.0 (5.0–6.0)	<0.001
Summary	Difficulty compared with reality	4.0 (3.0–4.0)	0.559

Overall score

The overall score for the simulation was statistically significant compared with a neutral score (median score 6.0, $P < 0.001$). When participants were asked to compare the level of difficulty of the simulator compared with real life the result was not significantly different (median score 4, $P = 0.559$).

Discussion

The simulation of practical procedures may be a valuable adjunct for medical training, particularly in the light of reduced working hours [28]. In colonoscopy, simulation is a growing field [3]. In recent years more research to support the use of *ex vivo* animal models has been published [4,29].

The results of this study have demonstrated that the WIMAT colonoscopy suitcase has excellent face and content validity across a range of parameters. A cohort of participants, experienced in the skill of colonoscopy and polypectomy, awarded the model favourable scores for visual, anatomical and mechanical realism. All of the measured parameters for visual realism scored well enough to produce a statistically significant result. Most encouragingly was the statistically significant score for the overall realism of the model and the non-statistically significant score comparing the difficulty of the simulation to actual reality. For anatomical realism, it was interesting to note that the scoring for simulated pedunculated polyps was more favourable than that for the simulated sessile polyps. However, when participants were asked to comment on the realism of performing a mucosal lift on a sessile polyp, a statistically significant favourable result was achieved. This would imply that the reduced level of anatomical realism of the sessile polyp did not significantly impact on the process of performing the polypectomy. The other non-statistically significant parameters were for 'resistance to scope movement' and 'paradoxical motion'. We would agree that this may in fact be a limitation of the current model. However, this should not significantly affect the use of the simulator which is designed to focus on polypectomy training as opposed to navigation and endoscopic steering.

There are several proposed benefits of using our novel *ex vivo* animal tissue polypectomy trainer. First, the cost of the simulator is considerably less than VR, live animal and cadaveric models. This makes it a financially viable option for training large numbers of participants with varying levels of experience. Also, the model is portable, has a simple set-up process which requires little technical expertise and can be tailored according to the level of experience of the user. This means that it can be utilized at any training centre with minimal inconvenience to faculty and course administrators. Limitations of this model are that it is single use and requires a time consuming process of polyp insertion. Furthermore, porcine bowel can also be relatively thin, risking perforation and desufflation during the simulation. We have overcome this by using rectal tissue which is much thicker than other parts of the porcine large intestine. The model can be quickly and easily patched should a perforation occur.

In conclusion, this paper highlights the benefits of *ex vivo* animal simulation and introduces our novel porcine simulator. We have confirmed the face and content validity of this model. Future work will focus on demonstrating its construct and concurrent validity and on testing the capacity of the model to allow transfer of skills into reality.

Acknowledgements

Neil Warren, Roger Leicester, Sunil Dolwani, Neil Hawkes, Stuart Goddard and Konstantinos Arnaoutakis designed and constructed the WIMAT endoscopy suitcase. James Ansell and Jared Torkington have no conflicts of interest or financial ties to disclose.

References

1 Markman HD. A new system for teaching proctosigmoidoscopic morphology. *Am J Gastroenterol* 1969; **52:** 65–9.
2 Classen M, Ruppin H. Practical training using a new gastrointestinal phantom. *Endoscopy* 1974; **6:** 127–31.
3 Desilets DJ, Banerjee S, Barth BA *et al.*, ASGE Technology Committee Endoscopic simulators. *Gastrointest Endosc* 2011; **73:** 861–7.
4 Sedlack RE, Baron TH, Downing SM, Schwartz AJ. Validation of a colonoscopy simulation model for skills assessment. *Am J Gastroenterol* 2007; **102:** 64–74.
5 The Chamberlain Group. *General and Gastrointestinal: Lower GI.* http://www.thecgroup.com/products/lower-gi (accessed 31 October 2011).
6 Simbionix. *GI Mentor.* http://simbionix.com/simulators/gi-bronch-gi-mentor/(accessed 31 October 2011).
7 WIMAT. *The WIMAT Colonoscopy Suitcase.* http://www.cardiff.ac.uk/pgmde/sections/wimat/simulationtrainingsupport/simulation-training-support.html (accessed 31 October 2011).
8 Fayez R, Feldman LS, Kaneva P, Fried GM. Testing the construct validity of the Simbionix GI Mentor II virtual reality colonoscopy simulator metrics: module matters. *Surg Endosc* 2010; **24:** 1060–5.
9 Felsher JJ, Olesevich M, Farres H *et al.* Validation of a flexible endoscopy simulator. *Am J Surg* 2005; **189:** 497–500.
10 Grantcharov TP, Carstensen L, Schulze S. Objective assessment of gastrointestinal endoscopy skills using a virtual reality simulator. *JSLS* 2005; **9:** 130–3.

11 Haycock AV, Bassett P, Bladen J, Thomas-Gibson S. Validation of the second-generation Olympus colonoscopy simulator for skills assessment. *Endoscopy* 2009; **41:** 952–8.
12 Koch AD, Haringsma J, Schoon EJ, de Man RA, Kuipers EJ. A second-generation virtual reality simulator for colonoscopy: validation and initial experience. *Endoscopy* 2008; **40:** 735–8.
13 Datta V, Mandalia M, Mackay S, Darzi A. The PreOp flexible sigmoidoscopy trainer. Validation and early evaluation of a virtual reality based system. *Surg Endosc* 2002; **16:** 1459–63.
14 MacDonald J, Ketchum J, Williams RG, Rogers LQ. A lay person versus a trained endoscopist: can the PreOp endoscopy simulator detect a difference? *Surg Endosc* 2003; **17:** 896–8.
15 Mahmood T, Darzi A. A study to validate the colonoscopy simulator. *Surg Endosc* 2003; **17:** 1583–9.
16 Sedlack RE, Kolars JC. Validation of a computer-based colonoscopy simulator. *Gastrointest Endosc* 2003; **57:** 214–8.
17 Hassan I, Osei-Agymang T, Radu D, Gerdes B, Rothmund M, Fernández ED. Simulation of laparoscopic surgery – four years experience at the Department of Surgery of the University Hospital Marburg. *Wien Klin Wochenschr* 2008; **120:** 70–6.
18 Maschuw K, Hassan I, Bartsch DK. Surgical training using simulator Virtual Reality. *Chirurg* 2010; **81:** 19–24.
19 Maagaard M, Sorensen JL, Oestergaard J *et al.* Retention of laparoscopic procedural skills acquired on a virtual-reality surgical trainer. *Surg Endosc* 2011; **25:** 722–7.
20 Sedlack RE. Validation process for new endoscopy teaching tools. *Tech Gastrointest Endosc* 2011; **13:** 151–4.
21 American College of Surgeons education division. *Technical Skills Education Surgery.* http://www.facs.org/education/technicalskills/faqs/faqs.html (accessed 31 October 2011).
22 Fresh Tissue Supplies. http://www.cdrex.com/fresh-tissue-supplies-limited-4468578.html (accessed 22 November 2011).
23 Europa. *Animal By-Products Not Intended for Human Consumption.* http://europa.eu/legislation_summaries/food_safety/animal_nutrition/f81001_en.htm (accessed 22 November 2011).
24 VELP® scientifica. *Heating Magnetic Stirrers.* http://www.velp.com/en/family/6/Stirring_line/26/Heating_magnetic_stirrers/ (accessed 22 November 2011).
25 Storm Case. http://pelican.com/ (accessed 22 November 2011).
26 Cook Medical. AcuSnare®. http://www.cookmedical.com/esc/dataSheet.do?id=5773 (accessed 22 November 2011).
27 Covidien. http://www.valleylab.com/product/es/generators/ez-over.html (accessed 22 November 2011).
28 Donohoe CL, Sayana MK, Kennedy MT, Niall DM. European working time directive: implications for surgical training. *Ir Med J* 2010; **103:** 57–9.
29 Hochberger J, Maiss J. Currently available *ex vivo* models. *Gastrointest Endoscopy Clin N Am* 2006; **16:** 435–49.

Editor's Choice

doi:10.1111/codi.12103

Surgeons are, by nature, eager to try new procedures and tools to improve patient care. This has not always been preceded by adequate training in properly validated teaching models. Ansell *et al.* address some of these issues in a paper on a new training model for interventional colonoscopy. In an *ex vivo* porcine model this group has elegantly shown that polypectomy may be taught efficiently because of the model's superior mucosal realism, endoscopic snare control and polyp handling. It may therefore be a more effective training model than virtual reality models. The paper also describes how teaching models should be validated and is a must-read if only for this reason.

Alexander Engel
Editor, *Colorectal Disease*

The Welsh Institute for Minimal Access Therapy colonoscopy suitcase has construct and concurrent validity for colonoscopic polypectomy skills training: a prospective, cross-sectional study

James Ansell, MBBCh,[1] Joanna J. Hurley, MBBCh,[2] James Horwood, MS,[3] Chantelle Rizan, BSc,[4] Konstantinos Arnaoutakis, MPhil,[1] Stuart Goddard, PGCE,[1] Neil Warren, PhD,[1] Jared Torkington, MS[3]

Cardiff, United Kingdom

Background: The Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase is an ex vivo porcine simulator for polypectomy training.

Objective: To establish whether this model has construct and concurrent validity.

Design: Prospective, cross-sectional study.

Setting: Endoscopic training center.

Participants: Twenty novice (N), 20 intermediate (I), 20 advanced (Ad), and 20 expert (E) colonoscopists.

Intervention: A simulated polypectomy task aimed at removing 2 polyps; A (simple), B (complex).

Main Outcome Measurements: Two accredited colonoscopists, blinded to group allocation, scored performances according to Direct Observation of Polypectomy Skills (DOPyS) assessment parameters. Group performances were compared. Real-life DOPyS scores were correlated to simulator DOPyS results.

Results: Median overall DOPyS scores for novices were 1.00 (1.00-1.87) for A and 0.50 (0.00-1.00) for B (A vs B; $P < .01$). Intermediates scored 2.50 (2.00-2.88) for A and 2.00 (1.13-2.50) for B (A vs B; $P = .03$). The advanced group scored 3.00 (2.50-3.50) for A and 2.50 (2.00-3.00) for B (A vs B; $P = .01$). Experts scored 3.00 (3.00-3.88) for A and 3.00 (2.50-3.50) for B (A vs B; $P = .47$). Intergroup comparisons for A were, N vs I; $P < .01$, N vs Ad; $P < .01$, N vs E; $P < .01$, I vs Ad; $P < .01$, I vs E; $P < .01$, and Ad vs E; $P = .46$. Intergroup comparisons for B were, N vs I; $P < .01$, N vs Ad; $P < .01$, N vs E; $P < .01$, I vs Ad; $P = .03$, I vs E; $P < .01$, and Ad vs E; $P = .06$. There was no difference between real-life DOPyS scores and simulator scores (0.07).

Limitations: The model does not have inbuilt assessment parameters.

Conclusion: This simulator demonstrates construct and concurrent validity for colon polypectomy training. (Gastrointest Endosc 2014;79:490-7.)

Colonoscopy training in the United Kingdom is governed by the Joint Advisory Group on Gastrointestinal Endoscopy (JAG).[1] Simulation forms an integral part of the JAG training process. There are several types of colonoscopy simulators available, ranging from bench models to virtual reality platforms.[2-6] The focus of these simulators

Abbreviations: DOPyS, Direct Observation of Polypectomy Skills; JAG, Joint Advisory Group on Gastrointestinal Endoscopy; WIMAT, Welsh Institute for Minimal Access Therapy.

DISCLOSURE: *J. Ansell was funded to conduct this research by the Royal College of Surgeons of England Research Fellowship scheme. All other authors disclosed no financial relationships relevant to this publication.*

0016-5107/$36.00
http://dx.doi.org/10.1016/j.gie.2013.08.003

Received February 11, 2013. Accepted August 4, 2013.

Current affiliations: Welsh Institute for Minimal Access Therapy (1), University Hospital Llandough, Cardiff (2), University Hospital of Wales, Wales (3), Cardiff University School of Medicine, Cardiff, Wales, United Kingdom (4).

Reprint requests: Mr James Ansell, MBBCh, Royal College of Surgeons Clinical Research Fellow in Simulation, Welsh Institute for Minimal Access Therapy (WIMAT), Cardiff Medicentre, Heath Park, Cardiff, CF14 4UJ, Wales, United Kingdom.

If you would like to chat with an author of this article, you may contact Mr Ansell at ansellj@cf.ac.uk.

is on endoscopic handling, which is useful for novice trainees. The benefit of these simulators is minimal once basic skills have been acquired.[7] Advanced trainees require simulation techniques that allow complex interventions such as polypectomy to be practiced in a safe environment.

The Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase is an ex vivo porcine simulator for colonoscopic polypectomy training.[8] This has the capacity to recreate a range of colonic polyps. A previous study has demonstrated that this simulator has good expert validity for skills training in polypectomy.[8] The next stage of any validation process is to evaluate construct and concurrent validity. Construct validity is the ability of a simulator to distinguish between differing levels of real-life expertise, and concurrent validity is the comparison of the simulator assessment to the criterion standard.[9]

Colonoscopic polypectomy is assessed in real life by using the Direct Observation of Polypectomy Skills (DOPyS) JAG form.[10] This is scored on a scale of 1 to 4 where 1 = *Accepted standards are not yet met, with frequent uncorrected errors;* 2 = *Some standards are not yet met, aspects to be improved, some uncorrected errors;* 3 = *Competent and safe throughout procedure, no uncorrected errors;* and 4 = *Highly skilled performance*. These scores are awarded against a list of 34 parameters that are divided into generic, stalked polyps, sessile polyps, and post-polypectomy categories. Polyps are defined as level 1 (<1 cm) and level 2 (>1 cm). United Kingdom trainees must demonstrate competency (>3 on DOPyS scoring) for level 1 polypectomy in order to achieve provisional accreditation.[11] Trainees can then apply for full accreditation when their scores are 3 to 4 on more than 90% of their last 4 DOPyS assessments for level 2 polypectomy.[11]

The aim of this study was to evaluate the construct and concurrent validity of the WIMAT colonoscopy suitcase by using the DOPyS assessment criteria relevant to stalked, level 1 polyps. A secondary aim was to test the hypothesis that the simulator can be reliably adjusted to recreate polypectomy tasks with varying degrees of difficulty.

Take-home Message

- This novel, ex vivo simulator can be used for training in simple polypectomy procedures. It is reproducible and inexpensive and can be constructed with minimal expertise.

METHOD

Study design

The Research Ethics Committee (REC) for Wales has stated that ethical approval is not required for this work under NHS research governance arrangements. A questionnaire was used to establish expertise and assign each participant to 1 of 4 groups. Novices had limited experience with colonoscopy (<50) but were pursuing careers in either gastroenterology or surgery. Intermediates were surgical or gastroenterology trainees between specialist trainee levels of 3 and 7. These participants all had the experience of performing <200 colonoscopies. This figure was chosen as the proposed number of colonoscopies needed to reach competency. Advanced participants were independent colonoscopists with full JAG accreditation. All advanced participants routinely perform colonoscopy and polypectomy as part of their standard practices. Experts were highly skilled bowel screening colonoscopists. All work in teaching hospitals with high-volume workloads. Previous endoscopic experience, simulator use, JAG courses attended or taught, and real-life DOPyS assessment scores were also recorded.

The simulation task

All participants followed an identical study protocol. After completing the questionnaire, a standardized instruction document explaining the task (including performing polypectomy) was read to each candidate, and a maximum of 5 minutes was allowed for participants to become familiar with the endoscopic equipment provided. No questions were permitted from any of the participants. The WIMAT colonoscopy suitcase was used to perform 2 standardized snare polypectomies on simulated level 1 pedunculated polyps (Fig. 1A and B). All polyps were 1 cm in size (from stalk base to polyp head). The length of the stalk exposed was 3 mm each time. Polyp A (simple) was inserted at 25 cm from the anal verge in the 6 o'clock position in front of a luminal fold. It was hypothesized that this position would allow participants to demonstrate a basic level of procedural skill. Polyp B (complex) was inserted at 43 cm from the anal verge and positioned at 1 o' clock, behind a luminal fold and distal to a simulated colon bend. This was designed to increase the difficulty of the polypectomy with the aim of differentiating between more experienced users. Every effort was made to ensure that the setup of each simulation was identical. The length of bowel used did not permit loop formation of the endoscope. Candidates were asked to insert the colonoscope and locate, snare, remove, and retrieve both polyps by using a Roth Net (US endoscopy, Mentor, Ohio). The same endoscopic assistant was present during all procedures. All polypectomies were video recorded for future analysis, and the files were coded in a random fashion by using a number between 1 and 80.

Performance evaluation

Two JAG-accredited colonoscopists, independent of the study design, analyzed each performance by using the DOPyS assessment form. Eight of 34 DOPyS assessment parameters were included according to their relevance to stalked polyps and generic skills. DOPyS parameters were excluded if they could not be assessed by video format or were not relevant to stalked polypectomy. Both assessors

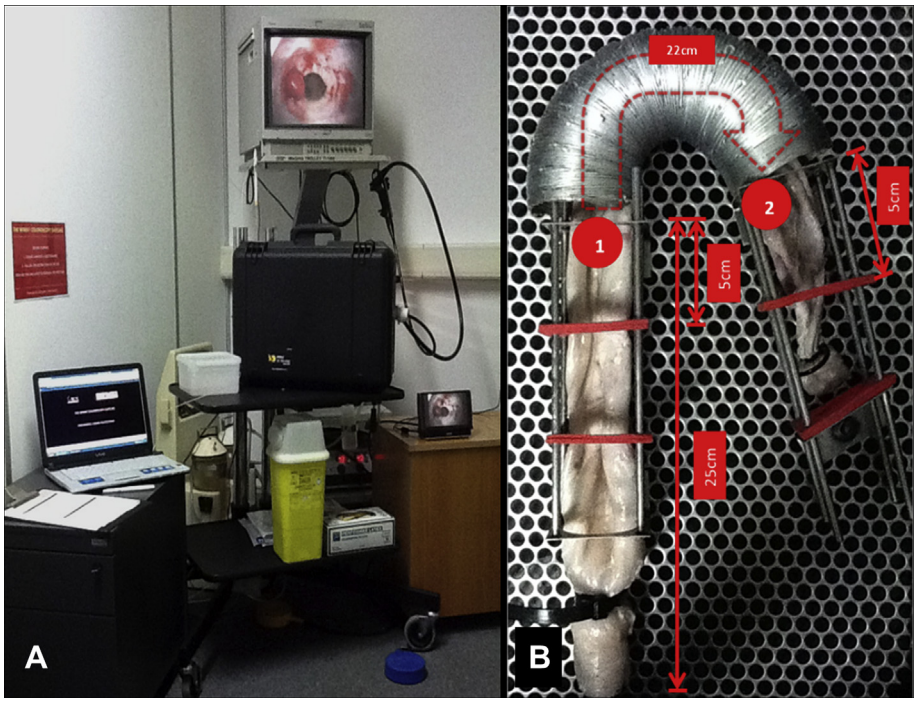


Figure 1. A, Overview of simulator set-up. **B,** Internal view of porcine bowel where 1 represents the position of polyp A, and 2 represents the position of polyp B.

were given DOPyS descriptor guides for generic and stalked polyps. The time taken to complete each polypectomy was recorded. Assessors remained blinded to the level of experience of the participant. On completion of the evaluation stage, all candidates were awarded a DOPyS score (from 1-4) for polyp A and polyp B, with their scores being compared between groups. Where available, real-life DOPyS scores were compared with simulator DOPyS scores.

Equipment

The same endoscopic equipment was used for all procedures. This included a Pentax EC-3840L colonoscope, Pentax PVM-2053MD monitor, and Pentax EPM-3500 processor (Pentax Medical, Slough, UK).[12] Standardized endoscopic snares (maximum diameter 25 mm) were supplied by Cook (Wilson-Cook Medical GI Endoscopy, Bloomington, Ind).[13] A 3.0-cm Roth Net was used for polyp retrieval.[14] The energy source used was the Valley Lab EZc (Covidien, Mansfield, Mass), placed on a coagulation setting of 35 W at all times.[15] An Eschmann VP25 portable suction unit (Eschmann, West Sussex, UK) set at a standard level was also used.[16] Performances were video recorded by using the Archos 7 MN 6700 (Archos, Southampton, UK) from the video output of the endoscopic processor.[17] There was no direct affiliation between this research and any of the medical equipment companies listed.

Data analysis

Previous research has demonstrated that experts (equivalent to our definition of experts) have an 88% (n = 15) chance of scoring 3 to 4 (or pass) on the DOPyS, whereas non-experts (equivalent to our definition of intermediate level) have a 53% (n = 8) chance of scoring between 3 and 4 (or pass) on the DOPyS.[10] Assuming that the novice group would take a similar drop in performance (from 53%-18%), then 20 participants in each group would give >80% power to detect a difference in performance of 35% between the groups, by using a 2-tailed test with a confidence level of 5%. A Wilcoxon signed rank test was used to assess differences between scores for polyps A and B within the same group and between simulator scores and real-life DOPyS scores. A Mann-Whitney *U* test was used to assess differences between scores for polyps A and B

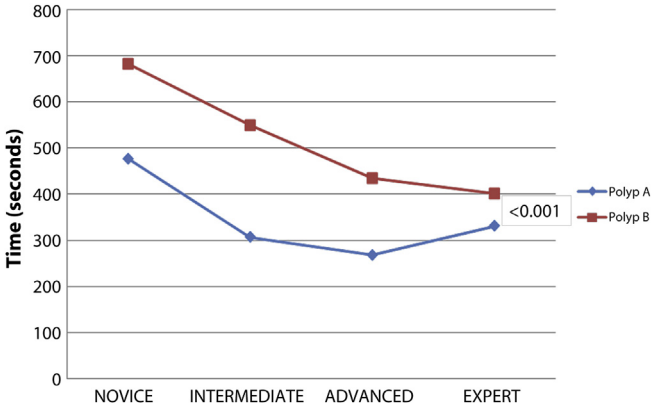


Figure 2. Median time taken (seconds) for each group for polyps A and B.

TABLE 1. Median (IQR) DOPyS scores for Polyp A and B for 2 independent JAG accredited reviewers*

Group	Polyp task	Direct observation of polypectomy skills parameter assessed		
		Attempts to achieve optimal position	Optimizes view by aspiration/insufflation/ wash	Directs snare accurately over polyp head
Novices (n = 20)	Polyp A	1.00 (1.00-1.87)	1.00 (1.00-1.50)	0.75 (0.00-1.50)
	Polyp B	1.00 (0.00-1.38)	1.00 (0.00-1.00)	1.00 (0.00-1.00)
	P value	< .01	< .01	.40
Intermediate (n = 20)	Polyp A	2.25 (2.00-2.50)	2.25 (2.00-2.50)	2.00 (2.00-2.88)
	Polyp B	1.75 (1.50-2.50)	2.00 (1.50-2.38)	1.75 (1.50-2.50)
	P value	< .01	< .01	.01
Advanced (n = 20)	Polyp A	3.00 (2.50-3.50)	3.00 (2.13-3.50)	3.00 (2.63-3.50)
	Polyp B	2.50 (2.00-3.00)	2.50 (2.00-3.00)	2.50 (2.00-3.00)
	P value	.01	.01	.04
Experts (n = 20)	Polyp A	3.25 (2.50-3.88)	3.00 (2.50-3.50)	3.00 (2.50-3.88)
	Polyp B	3.00 (2.50-3.50)	3.00 (2.50-3.38)	2.75 (2.50-3.50)
	P value	.07	.46	.31

*1 = Accepted standards are not yet met with frequent uncorrected errors; 2 = Some standards are not yet met, aspects to be improved, some uncorrected errors; 3 = Competent and safe throughout procedure, no uncorrected errors; and 4 = Highly skilled performance.

among groups. Multiple group comparisons were assessed by using the Kruskal-Wallis 1-way analysis of variance test. Data are expressed as medians with an interquartile range (IQR), where a *P* value of < .05 was considered statistically significant. Interrater reliability was compared by using the *k* statistic, where 0.81 to 1.00 indicates very good agreement, 0.61 to 0.80 good agreement, 0.41 to 0.60 moderate agreement, 0.21 to 0.40 fair agreement, and < 0.20 poor agreement. All calculations were performed on PASW Statistics 18 (10) (Hong Kong, China).

RESULTS

Eighty participants (20 per group) attempted the polypectomy task. All novices had previously performed <50 colonoscopies and no pedunculated polypectomies. Sixteen (80%) intermediates had performed <100 colonoscopies and <50 pedunculated polypectomies. In the advanced group, 15 participants (75%) had performed >500 colonoscopies and 14 (70%) had performed >50 pedunculated polypectomies. All experts reported performing >500 colonoscopies and 14 (70%) had performed >50 pedunculated polypectomies. The median number of times bench colonoscopy simulators were used by the cohort was 0.00 (1.00-3.00) (*P* < .01 between groups). For virtual reality and ex vivo colonoscopy simulators, this was 1.00 (0.00-3.00) (*P* < .01 between groups) and 0.00 (0.00-1.80) (*P* = .01 between groups), respectively. No novices had attended any JAG accredited colonoscopy courses whereas 12 (60%) intermediates reported completion of the JAG basic skills in colonoscopy course. Nine (45%) advanced participants and 16 (80%) experts have taught on the JAG advanced skills in colonoscopy course. Six (30%) advanced participants and 13 (65%) experts have also previously taught on the JAG polypectomy skills course.

Seven participants (8.8%) (all novices) failed to complete the polyp A task, and 7 participants (8.8%) (6 novices, 1 intermediate) failed to complete the polyp B task. Median (IQR) completion times (seconds) for polyp A were 477 (322-672) for novices, 307 (215-395) for intermediates, 268 (235-379) for the advanced group, and 331 (199-451) for experts (*P* = .06 between groups). For polyp B, the completion times were 683 (545-926) for novices, 550 (417-752) for intermediates, 434 (273-684) for the advanced group, and 401 (298-524) for experts (*P* < .01 between groups), as shown in Figure 2.

Overall median DOPyS scores (Table 1) for novices were 1.00 (1.00-1.87) for polyp A and 0.50 (0.00-1.00) for polyp B (*P* < .01). Intermediate group overall DOPyS scores were 2.50 (2.00-2.88) for polyp A and 2.00 (1.13-2.50) for polyp B (*P* = .03). Advanced participants scored 3.00 (2.50-3.50) for polyp A and 2.50 (2.00-3.00) for polyp B (*P* = .01). Expert DOPyS scores were 3.00 (3.00-3.88) for polyp A and 3.00 (2.50-3.50) for polyp B (*P* = .47). In the novice group, there was a general trend that scores

TABLE 1. Continued

Direct observation of polypectomy skills parameter assessed				
Places snare at appropriate position on stalk	**Mobilizes polyp to ensure appropriate amount of trapped tissue**	**Applies appropriate degree of diathermy**	**Retrieves/attempts retrieval of polyp**	**Overall competency**
1.00 (1.00-1.50)	1.25 (1.00-1.50)	1.00 (1.00-1.50)	1.00 (0.50-1.50)	1.00 (1.00-1.87)
0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)	0.50 (0.00-1.00)
< .01	< .01	< .01	< .01	< .01
2.50 (2.00-2.88)	2.50 (2.00-3.00)	2.50 (2.00-2.88)	2.50 (2.50-3.00)	2.50 (2.00-2.88)
2.00 (1.13-2.50)	2.00 (1.50-2.50)	2.00 (1.13-2.50)	2.00 (1.00-2.50)	2.00 (1.13-2.50)
.06	.01	.04	< .01	.03
3.00 (2.00-3.50)	3.00 (2.50-3.50)	3.00 (2.50-3.50)	3.00 (2.13-3.50)	3.00 (2.50-3.50)
2.75 (2.00-3.38)	2.50 (2.00-3.50)	2.75 (2.00-3.38)	2.25 (1.50-3.00)	2.50 (2.00-3.00)
.33	.01	.02	.01	.01
3.00 (2.50-3.88)	3.00 (3.00-3.88)	3.00 (2.50-3.88)	3.25 (2.63-3.50)	3.00 (3.00-3.88)
3.00 (2.50-3.50)	3.00 (2.50-3.50)	3.00 (2.13-3.50)	2.50 (2.50-3.50)	3.00 (2.50-3.50)
.69	.18	.64	.16	.47

decreased from polyp A to polyp B in a statistically significant manner for all DOPyS parameters except for the "*Directs snare accurately over polyp head*" variable ($P = .40$). This was the same for the intermediate and advanced groups excluding the "*Places the snare at appropriate position on the stalk*" variable ($P = .06$ and $P = .33$, respectively). No statistically significant difference between scores for polyp A and B for any parameter were exhibited in the expert group.

Differences in DOPyS parameter scores between the groups were statistically significant in the majority of cases except for when the advanced group was compared with the experts (Table 2). Results were highly significant when novices were compared with either intermediate, advanced, or expert participants for either polyp A or B. This was similarly the case when intermediate-group DOPyS scores were compared with those of the advanced or expert groups. Non-statistically significant scores were found for the "*Retrieves or attempts retrieval of polyp*" variable (intermediate vs advanced for polyp A; $P = .09$ and intermediate vs advanced polyp A; $P = .22$). No statistical difference was demonstrated for any parameter in either polypectomy between the advanced and expert groups (overall advanced vs expert polyp A; $P = .46$ and polyp B; $P = .06$).

Fifteen participants (75%) in the intermediate group submitted evidence of real-life DOPyS scores for level 1 pedunculated polypectomy. Median real-life DOPyS scores were comparable with the median simulator scores (Table 3). Five of 8 parameters showed no statistical difference when comparing real versus simulator DOPyS scores ("*Directs snare accurately over polyp head*" [$P = .13$], "*Places the snare at appropriate position on the stalk*" [$P = .09$], "*Mobilizes polyp to ensure appropriate amount of tissue is trapped*" [$P = .06$], "*Applies appropriate degree of diathermy*" [$P = .07$], and "*Overall competency at polypectomy*" [$P = .07$]). DOPyS parameters that were statistically significant included "*Attempts to achieve optimal position*" ($P = .02$), "*Optimizes view by aspiration/insufflation/wash*" (0.04), and "*Retrieves or attempts retrieval of polyp*" (0.03).

Analysis of interrater reliability of the 2 JAG assessors showed moderate to very good agreement for overall DOPyS parameters for novices (polyp A $k = 0.8$ and polyp B $k = 0.5$). There was fair agreement for intermediate group scores (polyp A $k = 0.4$ and polyp B $k = 0.4$) and advanced group scores (polyp A $k = 0.4$ and polyp B $k = 0.2$). In the expert cohort, this ranged from fair to moderate (polyp A $k = 0.6$ and polyp B $k = 0.2$). There was reduced agreement between assessors when the more technically demanding polyp B was scored.

DISCUSSION

This study demonstrates that the WIMAT colonoscopy suitcase has good construct and some evidence of concurrent validity for skills training in colonoscopic

TABLE 2. Differences for DOPyS scores between groups for polyp A and B*

		Direct observation of polypectomy skills parameter assessed							
Group	**Polyp task**	**Attempts to achieve optimal position**	**Optimizes view by aspiration/ insufflation/ wash**	**Directs snare accurately over polyp head**	**Places snare at appropriate position on stalk**	**Mobilizes polyp to ensure appropriate amount of trapped tissue**	**Applies appropriate degree of diathermy**	**Retrieves/ attempts retrieval of polyp**	**Overall competency**
N vs I	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
N vs A	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
N vs E	Polyp A	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
	Polyp B	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01	<0.01
I vs A	Polyp A	<0.01	0.01	0.01	0.03	0.03	0.02	**0.09**	0.01
	Polyp B	0.01	<0.01	0.01	0.02	0.04	0.04	**0.22**	0.03
I vs E	Polyp A	<0.01	0.01	<0.01	0.01	<0.01	<0.01	0.03	<0.01
	Polyp B	<0.01	<0.01	<0.01	0.01	<0.01	<0.01	<0.01	<0.01
A vs E	Polyp A	**0.289**	**0.91**	**0.86**	**0.50**	**0.50**	**0.74**	**0.95**	**0.46**
	Polyp B	**0.086**	**0.15**	**0.11**	**0.30**	**0.24**	**0.14**	**0.08**	**0.06**

N, Novice; *I,* intermediate; *A,* advanced; *E,* expert.
*Calculated using Mann-Whitney U (2 samples) test for non-parametric data; non-statistically significant results highlighted in bold.

polypectomy. DOPyS scores obtained on the simulator correlate with the real-life level of expertise of the user. Novice colonoscopists perform snare polypectomy on the simulator at an expected level of competency for their skills set. DOPyS scores improve on the simulator as the level of real-life experience increases. Real-life DOPyS scores, where available, reflect the simulator scores obtained on the model.

We hypothesized that placing polyps in differing positions inside the simulator would provide candidates with varying degrees of difficulty. The results indicate that this may be the case. The proposed complex polyp B took longer for all groups to remove and generally resulted in lower DOPyS scores being obtained. No statistical difference was demonstrated between scores for polyp A and B in the expert group. However, expert scores were lower for polyp B coupled with an increase in the time taken to perform the task. Reduced interrater agreement for polyp B DOPyS scores was also demonstrated. This correlated with increasing candidate real-life experience and may be because the DOPyS is not designed for assessing expert colonoscopists performing technically demanding polypectomy.

Intergroup comparison demonstrated statistically significant differences across a range of DOPyS parameters. This was particularly apparent when novices were compared with all other groups and in the majority of parameters when intermediate participants were compared with either advanced or expert participants. Although no statistically significant difference was found when we compared advanced and expert groups, further refinements in polyp placement and complexity may allow this to be detected if a difference does in fact exist. The most consistently nonsignificant DOPyS parameter when groups were compared was for "*Retrieves or attempts retrieval of polyp*." This may be explained by the way in which the simulated polyps are constructed. Polyps are made from a gelatin mix that is injected into porcine sausage skin, forming the polyp stalk. If the snare is not placed around the stalk, then the diathermy cuts through the mix and liquefies it. This causes the polyp to disintegrate, making retrieval difficult.

Comparison of simulator DOPyS scores with real-life scores showed no statistical difference for the majority of parameters, including overall competency. This provides evidence for the concurrent validity of this model. Statistical differences in favor of real-life DOPyS scores were demonstrated for optimal positioning and optimizing view. This could be a reflection of the model's ability to replicate real-life colonoscope handling or may be related to the difficulty in correlating retrospective real-life polypectomy tasks with the standardized simulated polypectomy being performed.

TABLE 3. Comparison of simulator DOPyS scores versus real-life DOPyS score for pedunculated polypectomy*

	DOPyS parameter assessed for intermediate group real-life DOPyS (n = 15)							
Polyp task	**Attempts to achieve optimal position**	**Optimizes view by aspiration/ insufflation/ wash**	**Directs snare accurately over polyp head**	**Places snare at appropriate position on stalk**	**Mobilizes polyp to ensure appropriate amount of trapped tissue**	**Applies appropriate degree of diathermy**	**Retrieves/ attempts retrieval of polyp**	**Overall competency**
Simulator DOPyS	1.75 (1.75-2.75)	1.75 (1.50-1.75)	2.00 (1.00-2.00)	2.00 (1.50-2.50)	2.00 (1.5-2.50)	2.00 (1.50-2.50)	2.00 (1.00-2.00)	2.00 (1.50-2.50)
Real-life DOPyS	2.00 (1.00-2.00)	2.00 (1.00-2.00)	2.00 (1.50-2.00)	2.00 (1.00-2.00)	2.00 (1.00-2.00)	1.00 (1.00-2.00)	2.00 (1.75-2.50)	2.00 (1.00-2.00)
P value	**.02**	**.04**	.13	.09	.06	.07	**.03**	.07

IQR, Interquartile range; *DOPyS,* direct observation of polypectomy skills.
*Wilcoxon signed-rank test for nonparametric data; statistically significant results highlighted in bold.

One limitation of this study is that it focused on simple snare polypectomy. Although this allows reliable standardization of the polypectomy tasks, it does not fully validate the model as useful for training in other therapeutic colonoscopic procedures. Further work is required if the model is to be used as a training platform for more complex interventions such as EMR and endoscopic submucosal dissection. A further limitation may be that some of the participants had reported previous experience of using ex-vivo colonoscopy simulation which may have influenced their overall performance. Although this number was low, a significant difference was found across the groups.

Although this model has the benefit of being realistic, portable, and cost-effective, it is limited by the fact that it does not have inbuilt parameters for assessment. Therefore, retrospective DOPyS analysis of videoed performances was used. This limited the number of DOPyS parameters that could be applied. If this model is to be fully utilized in training and assessment, it may be more useful to establish more quantitative forms of performance evaluation. Motion analysis has been used in laparoscopic surgery as an objective method of assessing ability.[18] Obstein et al[19] used a similar approach of kinematics to generate a score for colonoscopy technical skill. This has the potential to be applied to our polypectomy simulator to improve its capacity for assessment and accreditation.

In conclusion, the WIMAT colonoscopy suitcase demonstrates good construct and some evidence of concurrent validity for polypectomy training. However, it is important that the validation process is completed. Therefore, future research will be directed toward evaluating whether training on the simulator leads to skills transfer to the real-life setting.

REFERENCES

1. JAG (2012). Joint Advisory Group on Gastrointestinal Endoscopy. Available at: http://www.thejag.org.uk/TrainingforEndoscopists.aspx. Accessed August 1, 2012.
2. Nelson DB, Bosco JJ, Curtis WD, et al. Technology status evaluation report: endoscopy simulators. Gastrointest Endosc 2000;51:790-2.
3. Desilets DJ, Banerjee S, Barth BA, et al. Endoscopic simulators. Gastrointest Endosc 2011;73:861-7.
4. Hochberger J, Matthes K, Maiss J, et al. Training with the compact EASIE biologic endoscopy simulator significantly improves hemostatic technical skill of gastroenterology fellows: a randomized controlled comparison with clinical endoscopy training alone. Gastrointest Endosc 2005;61:204-15.
5. Long V, Kallo AN. AccuTouch Endoscopy Simulator: development, applications and early experience. Gastrointest Endosc Clin N Am 2006;16:479-87.
6. Kim WS, Woo HS, Ahn W, et al. Non-clinical evaluation of the KAIST-Ewha Colonoscopy Simulator II. Stud Health Tech Informat 2007;125:214-6.
7. Sedlack RE, Baron TH, Downing SM, et al. Validation of a colonoscopy simulation model for skills assessment. Am J Gastroenterol 2007;102:64-74.
8. Ansell J, Arnaoutakis K, Goddard S, et al. The WIMAT colonoscopy suitcase model: a novel porcine polypectomy trainer. Colorectal Dis. Epub 2012 June 2.
9. Sedlack RE. Training to competency in colonoscopy: assessing and defining competency standards. Gastrointest Endosc 2011;74:355-66; e351-52.
10. Gupta S, Anderson J, Bhandari P, et al. Development and validation of a novel method for assessing competency in polypectomy: direct observation of polypectomy skills. Gastrointest Endosc 2011;73: 1232-9.e2.
11. JAG trainee certification process—colonoscopy. http://www.thejag.org.uk/downloads%5CCertification%20Process%5CJAG%20Colonoscopy%20March%202011.pdf. Accessed August 1, 2012.
12. Pentax Colonoscopy (2013). http://www.pentax.co.uk/en/LifeCare_Colonoscopy.html. Accessed August 1, 2012.
13. Cook Medical (2013). Endoscopy. http://www.cookmedical.com/esc/dataSheet.do?id=5773Roth. Accessed January 8, 2013.
14. US endoscopy (2013). Roth net tissue retrivers. http://www.usendoscopy.com/endoscopy/roth-net-retrieval-devices/roth-net-foreign-body-retrievers.aspx. Accessed January 8, 2013.

15. Covidien (2013). Valley lab electrosurgical generators. http://www.valleylab.com/product/es/generators/index.html. Accessed January 8, 2013.
16. Eschmann (2013). VP25. http://www.eschmann.co.uk/operating-theatre/surgical-suction/show/vp25. Accessed January 8, 2013.
17. Archos (2013). http://www.archos.com/. Accessed January 8, 2013.
18. Smith SGT, Torkington J, Brown TJ, et al. A tool for assessing laparoscopic dexterity in the performance of a laboratory-based laparoscopic cholecystectomy. Surg Endosc 2002;16:640-5.
19. Obstein KL, Patil VD, Jayender J, et al. Evaluation of colonoscopy technical skill levels by use of an objective kinematic-based system. Gastrointest Endosc 2011;73:315-21.

The American Journal of Surgery (2014) 207, 32-38

The American Journal of Surgery®

Clinical Science

Can endoscopists accurately self-assess performance during simulated colonoscopic polypectomy? A prospective, cross-sectional study

James Ansell, B.Sc., M.B.B.Ch.[a,*], Joanna J. Hurley, M.B.B.Ch.[b], James Horwood, M.S.[a], Chantelle Rizan, B.Sc.[c], Konstantinos Arnaoutakis, M.Phil.[a], Stuart Goddard, B.Sc.[a], Neil Warren, Ph.D.[a], Jared Torkington, M.S.[d]

[a]*Welsh Institute for Minimal Access Therapy, Cardiff CF14 4UJ, UK;* [b]*University Hospital Llandough, Cardiff, UK;* [c]*Cardiff University, School of Medicine, Cardiff, UK;* [d]*University Hospital of Wales, Cardiff, UK*

KEYWORDS:
Simulation;
Endoscopy;
Polypectomy;
Training;
Self-assessment

Abstract

BACKGROUND: The aim of this study was to establish if endoscopists can reliably self-assess their ability to perform simulated colonic polypectomy.

METHODS: Novices, intermediates, advanced, and experts performed a video-recorded polypectomy task using the Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase simulator. This involved removal of a simple polyp (A) and a complex polyp (B). Participants self-assessed themselves using a Direct Observation of Polypectomy Skills (DOPyS) assessment form. Two blinded, independent, Joint Advisory Group on Gastrointestinal Endoscopy (JAG) accredited assessors graded each performance using the same DOPyS scoring. The Spearman coefficient was used to determine the correlation between self and assessors' scores.

RESULTS: Eighty participants completed the task. There was a weak correlation between assessors' scores and self-assessment scores for all groups (novices: $\rho = -.44$, $P = .85$; intermediates: $\rho = -.16$, $P = .51$; advanced: $\rho = .16$, $P = .50$; and experts: $\rho = .07$, $P = .76$). There was a strong correlation between scores from assessor 1 and 2 for polyp A ($\rho = .80$, $P \leq .01$) and polyp B ($\rho = .80$, $P \leq .01$).

CONCLUSIONS: The correlation between self-assessment and assessors' scores is weak. Novices and intermediates underestimate performance, whereas advanced and experts overestimate performance. Regular feedback may improve accuracy.

The ability to accurately self-assess performance is an important component of medical education.[1] This is particularly true in practical specialties such as gastroenterology and surgery. In the United Kingdom, the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) recommends that trainees use a personal development plan in order to highlight learning needs.[2] Part of this process requires a degree of self-assessment and trainer-based formative review. This can be used to identify strengths for development and highlight weaknesses for correction.[3]

Despite the proposed benefit, several studies from a variety of disciplines consistently indicate that the accuracy of self-assessment is poor.[3–6] The majority of research focuses on written and clinical knowledge rather than technical skills. The evidence is contradictory when practical

The authors declare no conflict of interest.

* Corresponding author. Tel.: +44-2929-682138; fax: +44-2920-682134.
E-mail address: ansellj@cf.ac.uk

Manuscript received May 13, 2013; revised manuscript May 22, 2013

0002-9610/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2013.05.019

tasks are considered. Moorthy et al[7] found that senior surgical trainees are accurate in their self-assessment of technical procedures in a simulated operating room, whereas Pandey et al[8] showed a poor correlation between self-appraisal and expert appraisal. Other articles report that high self-belief does not predict success, and in novices it corresponds negatively with skill.[9]

The benefit of self-assessment has yet to be determined for therapeutic colonoscopy. Technically difficult tasks such as polypectomy, endoscopic mucosal resection, and endoscopic submucosal dissection are increasing in response to early polyp detection from bowel screening programs.[10] The aim of this study was to determine if endoscopists of differing levels of expertise can accurately self-assess their performance in polypectomy using a novel simulator, the Welsh Institute for Minimal Access Therapy (WIMAT) colonoscopy suitcase. This is an ex vivo, porcine simulator validated for colonic polypectomy skills training.[11]

Methods

Participants

Eighty participants were recruited from the Wales Deanery, UK. A questionnaire established previous expertise and allowed each candidate to be assigned to 1 of the following groups: novices: junior doctors with no experience of colonoscopy, intermediates: specialist trainees (specialty training levels 3 to 7) with experience of performing more than 200 colonoscopies, advanced: JAG-accredited, independent colonoscopists; and experts: JAG-accredited bowel screening colonoscopists. For JAG accreditation (full certification), a trainee must be competent across several performance indicators including demonstrating a cecal intubation rate of more than 90% and a serious complication rate of less than .5%.[2] To become a bowel screening colonoscopist, an independent practitioner must pass an additional written and practical assessment.[2]

Study setting

All procedures took place within the simulated endoscopy suite at the WIMAT using the WIMAT colonoscopy suitcase simulator. The same endoscopic equipment was used for all procedures (Pentax, Slough, UK). Endoscopic snares (25 mm) (Cook Wilson Cook Medical GI Endoscopy, Bloomington, IN) and 3.0-cm Roth Nets (US Endoscopy, Mentor, OH) were used for polypectomy and retrieval. The energy source was (Valley Lab EZc, Covidien, CO) set at 35 W for coagulation for each task, and a portable suction unit (Eschmann VP25, West Sussex, UK) remained at the same pressure each time (200 mm Hg). All performances were recorded from the video output of the endoscopic processor using the Archos 7 MN 6700 (Archos, Greenwood Village, CO).

Study procedure

The WIMAT colonoscopy suitcase was used to perform 2 snare polypectomy procedures. This is a previously validated task, and the exact positioning of the polyps and setup of the simulator is described in earlier work (Fig. 1).[12] The task was designed to allow trainees to perform a simple polypectomy (polyp A) and a complex polypectomy (polyp B). The setup of each simulation was identical, and the same endoscopy assistant was present for all cases. The luminal view of all procedures was video recorded for future performance analysis.

Assessment

After the polypectomy task, each participant completed a modified self-assessed Direct Observation of Polypectomy Skills (DOPyS) JAG form.[13] This is marked on a scale of 1 to 4 as follows: 1: accepted standards are not yet met with frequent uncorrected errors; 2: some standards are not yet met, aspects to be improved, and some uncorrected errors, 3: competent and safe throughout procedure and no uncorrected errors; and 4: highly skilled performance. Any incomplete parameters are awarded a score of 0. Eight of 34 DOPyS parameters were selected according to their relevance to generic skills and for stalked polypectomy. DOPyS parameters were excluded if they could not be assessed by video format.

Two JAG-accredited colonoscopists reviewed each video and scored the performance using the same DOPyS assessment. Both assessors were given DOPyS descriptor guides for generic and stalked polyps. Assessors remained blinded to the level of experience of the participant at all times.

Statistical analyses

Previous research has shown that experts (equivalent to our definition of experts) have an 88% ($n = 15$) chance of scoring 3 to 4 (or pass) on the DOPyS, whereas nonexperts (equivalent to our definition of intermediate level) have a 53% ($n = 8$) chance of scoring between 3 and 4 (or pass) on the DOPyS.[12] Assuming that the novice group would take a similar drop in performance (from 53% to 18%), then 20 participants in each group would give a greater than 80% power to detect a difference in performance of 35% between the groups using a 2-tailed test with a confidence level of 5%.[12]

All statistical analyses were performed using the PASW Statistics Package 18, Hong Kong, China. An assessor's score for each of the 8 DOPyS parameters was calculated by averaging the score of polyp A and B across both assessors. Inter-rater reliability of assessors and the relationship between self-assessment and expert assessment were determined using Spearman ρ correlation coefficients where less than .30 was considered to be a weak correlation, between .30

and .50 was considered a moderate correlation, and greater than .50 was considered a strong correlation.[14] To complement this analysis, a Wilcoxon signed rank test was used to assess differences in median group performance. Data are expressed as medians with an interquartile range (IQR); a P value less than .05 was considered statistically significant.

Results

Eighty participants (20 per group) attempted the polypectomy task, and all completed the self-assessment process after the simulation. Seven (8.8%) novices failed to complete polypectomy A, and 6 (7.5%) novices and 1 (1.3%) intermediate failed to complete polypectomy B. A statistically significant difference was shown between the groups for median overall competency assessment scores (novices vs intermediates, $P < .01$; novices vs advanced, $P \leq .01$; novices vs experts, $P \leq .01$; intermediates vs advanced, $P = .02$; and intermediates vs experts, $P \leq .01$). No statistically significant difference was observed between advanced vs experts, $P = .14$. Significant differences were also demonstrated in the self-assessment median group scores for overall competency (novices vs intermediates, $P \leq .01$; novices vs advanced, $P \leq .01$; novices vs experts, $P \leq .01$; intermediates vs advanced, $P \leq .01$; intermediates vs experts, $P \leq .01$; and advanced vs experts, $P \leq .01$).

When the assessors' scores were compared with the self-assessment scores of the total cohort of 80 participants, the correlation was strong ($\rho = .70$, $P \leq .01$, Fig. 2). However, when the groups were considered separately, the correlation was weak. The novice median assessment score for overall competency was 1.00 (IQR = .50 to −1.25) compared with a self-assessment score of .50 (IQR = .00 to 1.00) ($\rho = -.44$, $P = .85$). For intermediates, the overall competency assessors' score was 2.37 (1.75 to 2.50), and the self-assessment score was 2.00 (1.00 to 2.00) ($\rho = -.16$, $P = .51$). In the advanced group, the overall competency assessors' score was 2.75 (2.50 to 3.25), and the self-assessment score was 3.00 (3.00 to 3.00) ($\rho = .16$, $P = .50$). For the experts, the assessors' overall score was 3.00 (2.75 to 3.44) versus a median self-assessment score of 4.00 (3.00 to 4.00) ($\rho = .07$, $P = .76$). A similar correlation was noted for the other 7 DOPyS parameters measured (Table 1).

Inter-rater reliability between the assessors showed a moderate to strong correlation for overall competency in polypectomy (novices [polyp A: $\rho = .62$, $P \leq .01$ and polyp B: $\rho = .50$, $P = .03$], intermediates: [polyp A, $\rho = .30$, $P = .30$ and polyp B: $\rho = .60$, $P \leq .01$], advanced [polyp A: $\rho = .50$,

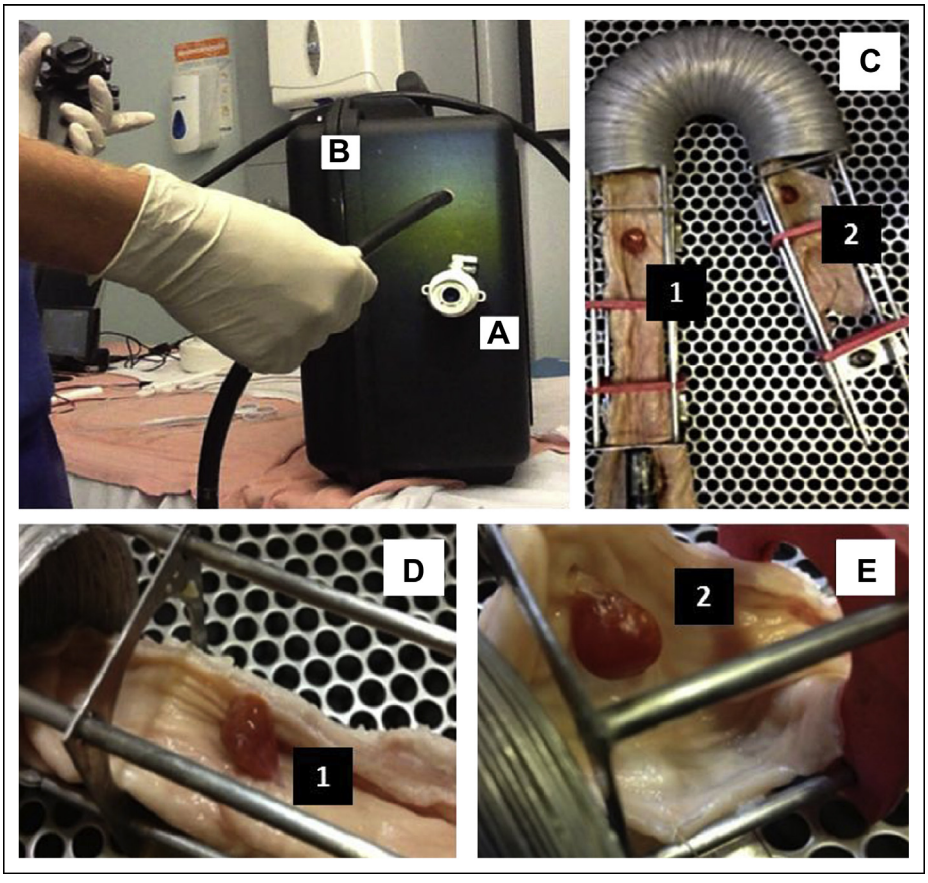


Figure 1 The WIMAT colonoscopy suitcase. (A) The laparoscopic entry port for the colonoscope, (B) the outer casing of the simulator, (C) the internal view of the simulator with the bowel opened to display the inserted polyps, (D) the position of polyp A (simple) inserted at 25 cm from the anal verge in the 6 o'clock position in front of a luminal fold, (E) the position of polyp B (complex) inserted at 43 cm from the anal verge and positioned at 1 o' clock behind a luminal fold and distal to a simulated colonic bend. 1 = polyp A; 2 = polyp B.

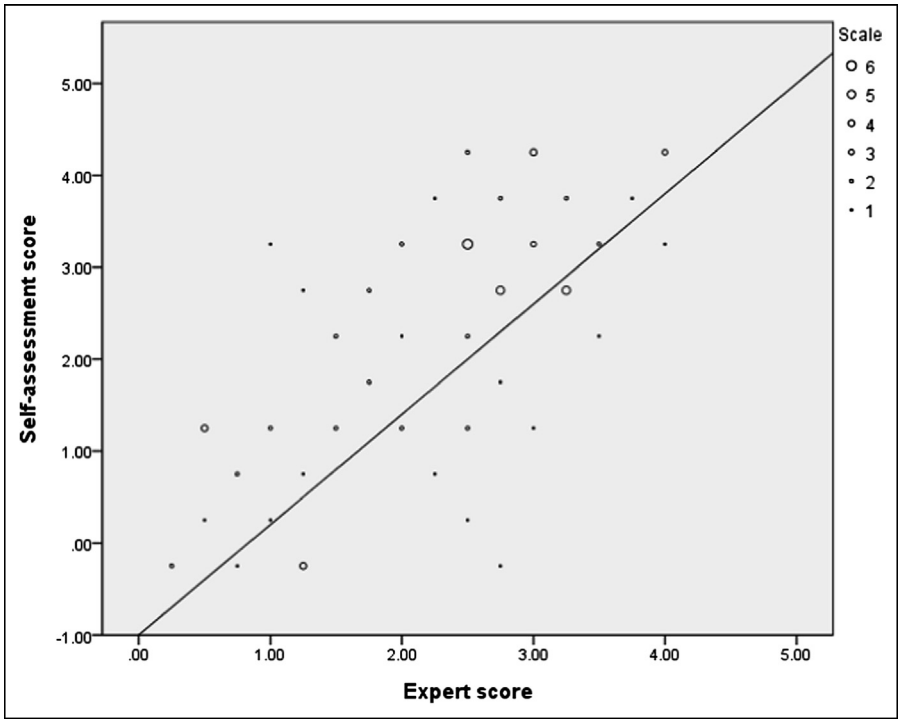


Figure 2 Expert (assessor's score) versus self-assessment of overall competency at polypectomy (binned data according to scale [right of graph]).

$P = .03$ and polyp B: $\rho = .62$, $P \leq .01$], and experts [polyp A: $\rho = .74$, $P \leq .01$ and polyp B: $\rho = .30$, $P = .21$). When an overall group comparison was performed for all groups together, the correlation was strong for polyp A ($\rho = .80$, $P \leq .01$, Fig. 3) and polyp B ($\rho = .80$, $P \leq .01$, Fig. 4).

Comments

This is the first study to evaluate the reliability of self-assessment during simulated colonoscopic polypectomy. There is a lack of consistency in the literature to conclude whether accurate self-assessment of technical skill is possible. A meta-analysis of 44 self-assessment studies in higher education reported a moderate correlation between self- and expert assessments of .39.[15] A similar review by Gordon[3] of 18 articles showed comparable findings. However, when medical trainee self-assessments are compared with expert scores, the correlation is usually weak.[8,16–18]

Our results for trainee self-assessment in simulated colonoscopic polypectomy are consistent with these findings. The relationship between self-assessment and independent assessment of advanced and expert colonoscopists again shows a weak statistical correlation. This is contrary to some reports in the literature that the ability to accurately self-assess improves with experience because the participant can recognize an expert performance and use this as a benchmark to assess his/her own skills.[7,19] Despite a weak statistical relationship, self-assessment scores did in general increase with real-life levels of expertise. Novices awarded themselves lower self-assessments scores than the intermediates and the intermediates lower scores than the advanced and experts.

Novices and intermediates tended to underestimate their ability (median assessors' score for novices = 1.00 [IQR .50 to 1.25] and 2.38 (IQR 1.75 to 2.50) for intermediates compared with a self-assessment median score for novices of .50 [IQR .00 to 1.00] and 2.00 [IQR 1.00 to 2.00] for intermediates). On the other hand, the advanced and experts groups overestimated their ability (median assessor score for advanced = 2.75 [IQR 2.50 to 3.25] and 3.00 (IQR 2.75 to 3.44) compared with self-assessment median scores of 3.0 [IQR 3.0 to 3.0] for advanced and 4.00 [IQR 3.00 to 4.00] for experts). These findings can be interpreted in several ways. First, trainees lack experience of performing real-life polypectomy and are therefore less able to accurately self-assess their performance. Advanced and expert colonoscopists may be familiar with assessing novice and intermediate performances but may be less able to repeat this process for themselves. They may also feel pressured to represent themselves in the best possible light. This term is known as impression management and has been previously applied to trainees but may, in fact, also be true of trainers themselves.[20]

Alternatively, the over- or underestimation can be directed at the assessment process itself. There are currently no valid, quantitative measures of polypectomy assessment. The DOPyS score is a subjective measure of skill that may be open to interpretation. It has been validated for the assessment of a range of real-life polypectomy procedures but has not been used in simulated polypectomy assessment previously. This may affect the reliability of its use in this trial. However, 2 blinded,

Table 1 Summary of assessors' scores versus self-assessment scores for all groups and for each DOPyS parameter used

		DOPyS parameter assessed							
Group	Score	Attempts to achieve optimal position	Optimizes view by aspiration/ insufflation/ wash	Directs snare accurately over polyp head	Places the snare at appropriate position on the stalk	Ensures appropriate amount of tissue trapped	Applies appropriate degree of diathermy	Retrieves or attempts for retrieval of polyp	Overall competency at polypectomy
Novices (n = 20)	Assessor score, median (IQR)	1.00 (.75–1.25)	1.00 (.50–1.25)	.63 (.63–1.25)	1.00 (.50–1.25)	1.00 (.56–1.19)	1.00 (.50–1.25)	1.00 (.31–1.25)	1.00 (.50–1.25)
	Self-assessment, median (IQR)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)	.50 (.00–1.00)
	Spearman ρ	−.71	−.10	.14	−.44	−.12	.53	−.07	−.44
	P value	*P* = .77	*P* = .68	*P* = .56	*P* = .85	*P* = .61	*P* = .83	*P* = .77	*P* = .85
Intermediate (n = 20)	Assessor score, median (IQR)	2.00 (1.75–2.69)	2.13 (1.75–2.50)	2.00 (1.75–2.50)	2.13 (1.50–2.50)	2.25 (1.75–2.75)	2.13 (1.81–2.50)	2.25 (2.00–2.75)	2.38 (1.75–2.50)
	Self-assessment Median (IQR)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)	2.00 (1.00–2.00)
	Spearman's (ρ)	.06	.01	−.20	−.14	−.11	−.23	−.15	−.16
	P value	*P* = .82	*P* = .97	*P* = .39	*P* = .56	*P* = .65	*P* = .33	*P* = .54	*P* = .51
Advanced (n = 20)	Assessor score Median (IQR)	2.87 (2.31–3.25)	2.75 (2.25–3.19)	2.75 (2.31–3.19)	2.75 (2.31–3.25)	3.00 (2.25–3.25)	2.75 (2.31–3.25)	2.50 (2.06–3.19)	2.75 (2.50–3.25)
	Self-assessment Median (IQR)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)	3.0 (3.0–3.0)
	Spearman's (ρ)	−.11	.09	.18	.19	.14	−.06	−.10	.16
	P value	*P* = .64	*P* = .72	*P* = .46	*P* = .42	*P* = .56	*P* = .80	*P* = .68	*P* = .50
Experts (n = 20)	Assessor score Median (IQR)	3.13 (2.56–3.69)	3.00 (2.50–3.44)	2.75 (2.50–3.50)	3.00 (2.56–3.44)	3.00 (2.56–3.44)	3.00 (2.56–3.44)	3.00 (2.56–3.44)	3.00 (2.75–3.44)
	Self-assessment Median (IQR)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)	4.00 (3.00–4.00)
	Spearman's (ρ)	.12	.16	.36	.39	.22	.27	.03	.07
	P value	*P* = .62	*P* = .50	*P* = .12	*P* = .09	*P* = .34	*P* = .26	*P* = .91	*P* = .76

Spearman ρ correlation: <.30 was considered to be a weak correlation, between .30 and .50 was considered a moderate correlation, and >.50 was considered a strong correlation.

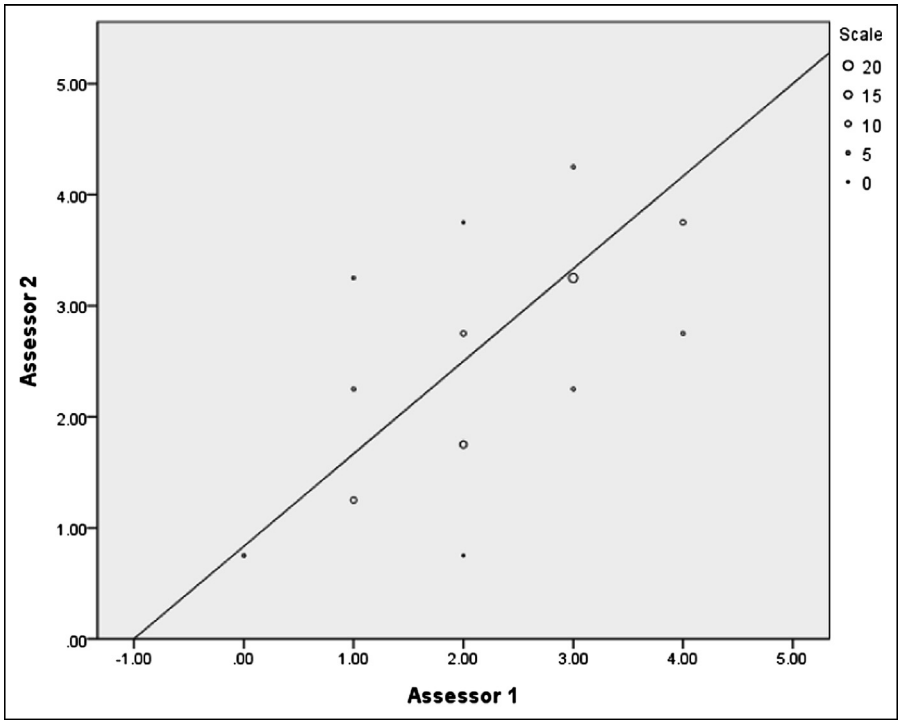


Figure 3 Assessor 1 versus assessor 2 scores for the polyp A task (binned data according to scale [right of graph]).

independent JAG-accredited assessors showed strong correlations for both polyp tasks, which implies that the assessor scores are in fact, accurate.

Finally, the process of self-assessment and independent assessment differed. Assessors watched video recordings retrospectively and therefore had time to scrutinize technique. Self-assessment scores were generated directly after the procedure. Martin et al[21] showed that the correlation between expert and self-assessment improves from .38 to .52 for communication skills if the participant reviews the video performance rather than from memory. This effect is diminished in more senior residents and experts.[21] It would be interesting to see if this effect was more pronounced in the self-assessment of technical skills. Different

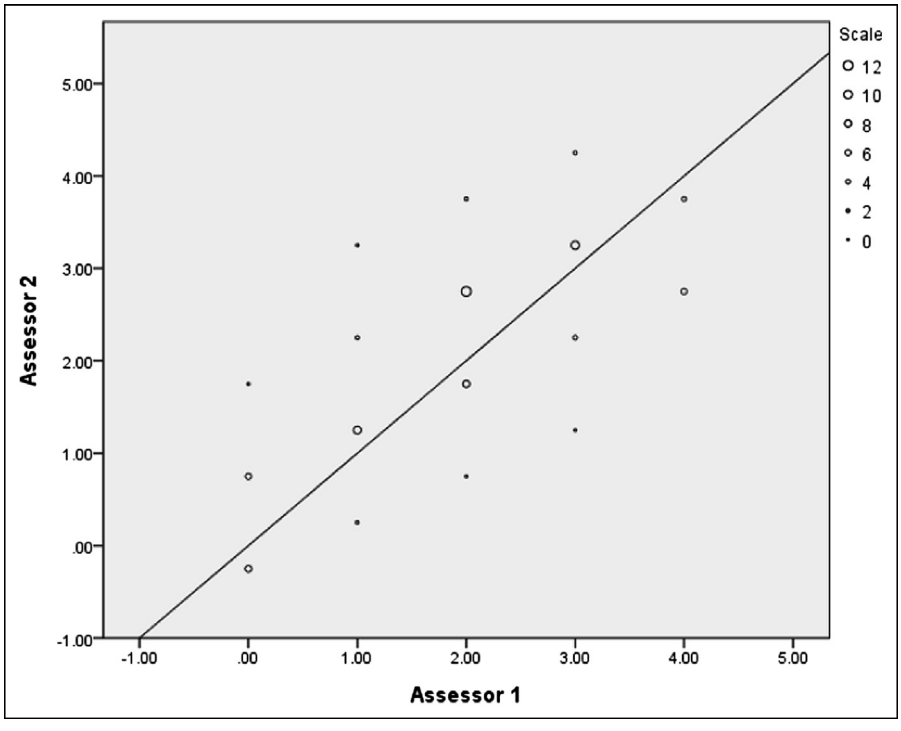


Figure 4 Assessor 1 versus assessor 2 scores for the polyp B task (binned data according to scale [right of graph]).

methods of conducting these appraisals should be explored in the future to ensure that self-assessment is utilized to its full potential in training and quality assurance.

A limitation of this trial may be that a simulator was used to replicate a real-life scenario. This was used to standardize the polypectomy task, which would be difficult to achieve in reality. This simulator has undergone rigorous validation studies previously but could never completely replicate real-life colonoscopic polypectomy. It may be interesting to assess the correlation between self-assessments and expert assessments using live polypectomy cases to see if the accuracy improves.

This research shows that there is a weak statistical correlation between self-assessment and independent expert assessment for simulated colonoscopic polypectomy. It remains to be seen if improvements in perceived performance may be achieved if participants self-assess video performance. Independent expert assessment seems to remain a strong way of reliably assessing performance in therapeutic colonoscopy.

References

1. Fitzgerald JT, Gruppen LD, White CB. The influence of task formats on the accuracy of medical students' self-assessments. Acad Med 2000;75:737–41.
2. JAG 2013. The Joint Advisory Group on GI Endoscopy (JAG). Available at: http://www.thejag.org.uk. Accessed April 2, 2013.
3. Gordon MJ. A review of the validity and accuracy of self-assessments in health professions training. Acad Med 1991;66:762–9.
4. Morton JB, Macbeth WAAG. Correlations between staff, peer and self-assessment of fourth year students in surgery. Med Educ 1997;11: 161–70.
5. Kegel-Flom P. Predicting supervisor, peer and self ratings of interim performance. J Med Educ 1975;50:812–5.
6. Stuart M, Goldstein R, Harris S, et al. Self-evaluation by residents in family medicine. J Fam Pract 1980;10:639–42.
7. Moorthy K, Munz Y, Adams S, et al. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. Am J Surg 2006;192:114–8.
8. Pandey VA, Wolfe JH, Black SA, et al. Self-assessment of technical skill in surgery: the need for expert feedback. Ann R Coll Surg Engl 2008;90:286–90.
9. Maschuw K, Osei-Agyemang T, Weyers P, et al. The impact of self-belief on laparoscopic performance of novices and experienced surgeons. World J Surg 2008;32:1911–6.
10. Towler BP, Irwig L, Glasziou P, et al. Screening for colorectal cancer using the faecal occult blood test, hemoccult. Cochrane Database Syst Rev 2000;2:CD001216.
11. Ansell J, Arnaoutakis K, Goddard S, et al. The WIMAT colonoscopy suitcase model: a novel porcine polypectomy trainer. Colorectal Dis 2013;15:217–23.
12. Ansell J, Hurley J, Horwood J, et al. The WIMAT colonoscopy suitcase has construct and concurrent validity for colonoscopic polypectomy skills training: a prospective, cross sectional study. Gastrointest Endosc; 2013 (In Press).
13. Gupta S, Bassett P, Man R, et al. Validation of a novel method for assessing competency in polypectomy. Gastrointest Endosc 2012;75:568–75.
14. Cohen J. Statistical Power Analysis for Behavioural Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum; 1988.
15. Falchikov NB, Boud D. Student self-assessment in higher education: a meta-analysis. Rev Educ Res 1989;59:395–430.
16. Evans AW, Leeson RM, Petrie A. Reliability of peer and self-assessment scores compared with trainers scores following third molar surgery. Med Educ 2007;41:866–72.
17. Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physicians self-assessment compared with observed measures of competence: a systematic review. JAMA 2006;296:1094–102.
18. Brewster LP, Risucci DA, Joehl RJ, et al. Comparisons of resident self-assessment with trained faculty and standised patient assessment of clinical and technical skills in a structured educational module. Am J Surg 2008;195:1–4.
19. Ward M, MacRae H, Schlachta C, et al. Resident self-assessment of operative performance. Am J Surg 2003;185:521–4.
20. Evans AW, Leeson RM, Newton-John TR. The influence of self-deception and impression management on surgeons' self-assessment scores. Med Educ 2002;36:1095.
21. Martin D, Regehr G, Hodges B, et al. Using videotaped benchmarks to improve the self-assessment ability of family practice residents. Acad Med 1998;73:1201–6.