A Study to Design an Objective Scoring System for Basic Surgical Skills (BSS)

Shazrinizam Shaharan
Royal College of Surgeons in Ireland, shazrinizamshaharan@rcsi.ie
A Study to Design an Objective Scoring System for Basic Surgical Skills (BSS)

Dr Shazrinizam Shaharan
MB BCh BAO BA MRCS
Department of Surgical Affairs
Royal College of Surgeons in Ireland

A thesis submitted to the School of Postgraduate Studies, Royal College of Surgeons in Ireland, in fulfillment of the degree of Medical Doctorate
April 2018

Supervised by
Professor Paul Neary MD FRCSI
Clinical Professor of Surgery, Trinity College Dublin

&
Professor Oscar Traynor MCh FRCSI
Professor of Postgraduate Surgical Education, RCSI

Co-Supervisor
Mr Donncha M Ryan BSc
Learning Technology Manager, RCSI
Candidate Thesis Declaration

I declare that this thesis, which I submit to RCSI for examination in consideration of the award of a higher degree (Medical Doctorate, MD) is my own personal effort. Where any of the content presented is the result of input or data from a related collaborative research programme this is duly acknowledged in the text such that it is possible to ascertain how much of the work is my own. I have not already obtained a degree in RCSI or elsewhere on the basis of this work. Furthermore, I took reasonable care to ensure that the work is original, and, to the best of my knowledge, does not breach copyright law, and has not been taken from other sources except where such work has been cited and acknowledged within the text.

Signed

Student Number 12140341

Date 30th April 2018
# Table of Contents

Candidate Thesis Declaration ........................................ ii

Table of Contents ....................................................... iii

List of Abbreviations .................................................. ix

List of Figures ......................................................... x

List of Tables ........................................................ xiv

Summary ..................................................................... xv

Acknowledgements ...................................................... xvi
## Table of Contents

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Introduction</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.1 Background of Study</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1.2 Context and Surgical Training</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>1.3 Technology and Surgical Training</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1.4 Observer Dependant Assessment Tools</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>1.5 Objectives</td>
<td>13</td>
</tr>
<tr>
<td>2</td>
<td>2.1 Ethical Consideration</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.2 Recruitment of Participants</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.3 Participant Demographics</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>2.4 Basic Surgical Skills Assessment</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>2.5 Assessment Tools</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2.5.1 Observational Assessment Tool</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2.5.2 Non-Observational Assessment Tool</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>2.6 Application of Metrics from Patriot™ Motion Tracking Device</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2.7 Statistical Analysis</td>
<td>27</td>
</tr>
<tr>
<td>3</td>
<td>3.1 Introduction</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Validity and Reliability of Global Rating Scale (GRS) in Open Surgical Skills</td>
<td>32</td>
</tr>
</tbody>
</table>
Table of Contents

3.2 Objectives 36

3.3 Materials and Methods 37
   3.3.1 Recruitment of Participants 37
   3.3.2 Participant Demographics 37
   3.3.3 Skills Assessment Tool 39
   3.3.4 Surgical Skills Assessment 39
   3.3.5 Statistical Analysis 40

3.4 Results 40
   3.4.1 Construct Validity of mGRS for One-handed Knot Tying Skill 40
   3.4.2 Construct Validity of mGRS for Simple Interrupted Suturing Skill 44
   3.4.3 Reliability of mGRS: One-handed Knot Tying Skill 51
   3.4.4 Reliability of mGRS: Simple Interrupted Suturing Skill 52
   3.4.5 Inter-rater Reliability: One-handed Knot Tying Skill 53
   3.4.6 Inter-rater reliability: Simple Interrupted Suturing Skill 55

3.5 Discussion 56

Chapter 4 Validity of a Motion Tracking System in Basic Surgical Skills 59
   4.1 Introduction 60
   4.2 Objectives 61
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2.1</td>
<td>Hypothesis Underlying the Objectives</td>
<td>61</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Detailed Objectives</td>
<td>62</td>
</tr>
<tr>
<td>4.3</td>
<td><strong>Materials and Methods</strong></td>
<td>62</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Recruitment of Participants</td>
<td>62</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Participants Demographics</td>
<td>63</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Skills Assessment Tools</td>
<td>63</td>
</tr>
<tr>
<td>4.3.4</td>
<td>Surgical Skills Assessment</td>
<td>64</td>
</tr>
<tr>
<td>4.3.5</td>
<td>Statistical Analysis</td>
<td>64</td>
</tr>
<tr>
<td>4.4</td>
<td><strong>Results</strong></td>
<td>65</td>
</tr>
<tr>
<td>4.4.1</td>
<td>Participant Demographics</td>
<td>65</td>
</tr>
<tr>
<td>4.4.2</td>
<td>Construct Validity of the Metrics in One-handed Knot Tying</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Skill</td>
<td></td>
</tr>
<tr>
<td>4.4.3</td>
<td>Construct Validity of the Metrics in Simple Interrupted Suturing Skill</td>
<td>74</td>
</tr>
<tr>
<td>4.4.4</td>
<td>Concurrent Validity of the Metrics in One-handed Knot Tying</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>Skill</td>
<td></td>
</tr>
<tr>
<td>4.4.5</td>
<td>Concurrent Validity of the Metrics in Simple Interrupted Suturing Skill</td>
<td>82</td>
</tr>
<tr>
<td>4.5</td>
<td><strong>Discussion</strong></td>
<td>84</td>
</tr>
<tr>
<td>Chapter 5</td>
<td>The application of Patriot™ motion tracking system in a proficiency-based assessment of basic surgical skills</td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>5.1 Introduction</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>5.2 Objectives</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>5.2.1 Hypothesis Underlying the Objectives</td>
<td>89</td>
<td></td>
</tr>
<tr>
<td>5.2.2 Detailed Objectives</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>5.3 Materials and Methods</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>5.3.1 Recruitment of Participants</td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>5.3.2 Surgical Skills Assessment</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5.3.3 Skills Assessment Tool</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5.3.4 Statistical Analysis</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>5.4 Results</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>5.4.1 Proficiency-based Assessment of Skills</td>
<td>93</td>
<td></td>
</tr>
<tr>
<td>5.4.2 Monitoring Trainees’ Progress</td>
<td>101</td>
<td></td>
</tr>
<tr>
<td>5.5 Discussion</td>
<td>111</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 6</th>
<th>Discussion &amp; Future Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 The Aims of Thesis</td>
<td>115</td>
</tr>
<tr>
<td>6.2 Summary of Main Findings</td>
<td>116</td>
</tr>
</tbody>
</table>
# Table of Contents

6.3 General Conclusions .............................................. 117

6.4 Future Works ....................................................... 121

Appendices ............................................................... 122

Appendix I Written Instruction for Each Task .................. 123

Appendix II Letter of Ethical Approval ............................ 124

Appendix III Subject Information Sheet ......................... 125

Appendix IV Consent Form for Participants ..................... 126

References ............................................................... 127

List of Publications and Presentations .......................... 135
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSATS</td>
<td>Objective structured assessment of technical skill</td>
</tr>
<tr>
<td>GRS</td>
<td>Global rating scale</td>
</tr>
<tr>
<td>VR</td>
<td>Virtual reality</td>
</tr>
<tr>
<td>SAGES</td>
<td>Society of American Gastrointestinal and Endoscopic Surgeons</td>
</tr>
<tr>
<td>FLS</td>
<td>Fundamentals of laparoscopic surgery</td>
</tr>
<tr>
<td>FES</td>
<td>Fundamentals of endoscopic surgery</td>
</tr>
<tr>
<td>MIS</td>
<td>Minimally invasive surgery</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective structures clinical examination</td>
</tr>
<tr>
<td>GOALS</td>
<td>Global operative assessment of laparoscopic skills</td>
</tr>
<tr>
<td>HST</td>
<td>Higher surgical training</td>
</tr>
<tr>
<td>IR</td>
<td>Inter-rater</td>
</tr>
<tr>
<td>RCSI</td>
<td>Royal College of Surgeons in Ireland</td>
</tr>
<tr>
<td>NSTC</td>
<td>National surgical training centre</td>
</tr>
<tr>
<td>BST</td>
<td>Basic surgical training</td>
</tr>
<tr>
<td>mGRS</td>
<td>Modified Global rating scale</td>
</tr>
<tr>
<td>SEU</td>
<td>System electronic unit</td>
</tr>
<tr>
<td>6DOF</td>
<td>Six degrees of freedom</td>
</tr>
<tr>
<td>TPL</td>
<td>Total path length</td>
</tr>
<tr>
<td>POI</td>
<td>Point of interest</td>
</tr>
<tr>
<td>ICSAD</td>
<td>Imperial college surgical assessment device</td>
</tr>
<tr>
<td>HUESAD</td>
<td>Hiroshima University Endoscopic Surgical Assessment Device</td>
</tr>
<tr>
<td>MISTELS</td>
<td>McGill Inanimate System for Training and Evaluation of Laparoscopic Skills</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A Sample of Task-specific Checklist from OSATS</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>The Global Rating Scale (GRS) of Operative Performance</td>
<td>11</td>
</tr>
<tr>
<td>1.3</td>
<td>Global Assessment of Laparoscopic Surgery (GOALS) Scoring System</td>
<td>12</td>
</tr>
<tr>
<td>2.1</td>
<td>The Original Global Rating Scale (GRS) in the Objective Structured Assessment</td>
<td>20</td>
</tr>
<tr>
<td>2.2</td>
<td>The Modified Global Rating Scale (mGRS) for Knot Tying Skill</td>
<td>21</td>
</tr>
<tr>
<td>2.3</td>
<td>A Screenshot of PiMgr Software</td>
<td>23</td>
</tr>
<tr>
<td>2.4</td>
<td>The Set-up for Knot Tying Skill Assessment with the Patriot™ Motion Tracking System</td>
<td>24</td>
</tr>
<tr>
<td>2.5</td>
<td>The Set-up for Simple Interrupted Suturing Assessment with the Patriot™ Motion Tracking System</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>The Patriot™ Sensor Attached to a Trainee’s Index Finger while Performing Simple Interrupted Suturing Task</td>
<td>26</td>
</tr>
<tr>
<td>2.7</td>
<td>Flow Diagram of the Experiment Process in Chapter 3 – Validity and Reliability of Global Rating Scale in Open Surgical Skills</td>
<td>29</td>
</tr>
<tr>
<td>2.8</td>
<td>Flow Diagram of the Experiment Process in Chapter 4 – Validity of a Motion Tracking System in Basic Surgical Skills</td>
<td>30</td>
</tr>
<tr>
<td>2.9</td>
<td>Flow Diagram of the Experiment Process in Chapter 5 – The Application of Patriot™ Motion Tracking System in a Proficiency-based Assessment of Basic Surgical Skills</td>
<td>31</td>
</tr>
<tr>
<td>3.1</td>
<td>Distribution of mGRS Scores for Item “Respect for Tissue” in One-handed Knot Tying Skill</td>
<td>41</td>
</tr>
<tr>
<td>3.2</td>
<td>Distribution of mGRS Scores for Item “Time and Motion” in One-handed Knot Tying Skill</td>
<td>42</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>3.3</td>
<td>Distribution of mGRS Scores for Item “Flow of Operation” in One-handed Knot Tying Skill</td>
<td>43</td>
</tr>
<tr>
<td>3.4</td>
<td>Distribution of mGRS Scores for Item “Respect for Tissue” in Simple Interrupted Suturing Skill</td>
<td>45</td>
</tr>
<tr>
<td>3.5</td>
<td>Distribution of mGRS Scores for Item “Time and Motion” in Simple Interrupted Suturing Skill</td>
<td>46</td>
</tr>
<tr>
<td>3.6</td>
<td>Distribution of mGRS Scores for Item “Instrument Handling” in Simple Interrupted Suturing Skill</td>
<td>47</td>
</tr>
<tr>
<td>3.7</td>
<td>Distribution of mGRS Scores for Item “Knowledge of Instrument” in Simple Interrupted Suturing Skill</td>
<td>48</td>
</tr>
<tr>
<td>3.8</td>
<td>Distribution of mGRS Scores for Item “Flow of Operation” in Simple Interrupted Suturing Skill</td>
<td>49</td>
</tr>
<tr>
<td>3.9</td>
<td>Distribution of mGRS Scores for Item “Knowledge of Procedure” in Simple Interrupted Suturing Skill</td>
<td>50</td>
</tr>
<tr>
<td>4.1</td>
<td>Number of Trainees who Participated According to Their Year of Training</td>
<td>67</td>
</tr>
<tr>
<td>4.2</td>
<td>Distribution of Average Time Taken to Complete One-handed Knot Tying Task</td>
<td>69</td>
</tr>
<tr>
<td>4.3</td>
<td>Distribution of the Total Path Length (TPL) between the Three Subject Groups</td>
<td>70</td>
</tr>
<tr>
<td>4.4</td>
<td>Distribution of the Average Distance from the Point of Interest (POI) or Centre of the Bench Model between the Three Subject Groups</td>
<td>71</td>
</tr>
<tr>
<td>4.5</td>
<td>Distribution of the Average Distance from the Z-axis Between the Three Subject Groups</td>
<td>72</td>
</tr>
<tr>
<td>4.6</td>
<td>Distribution of Average Time Taken to Complete Simple Interrupted Suturing Task</td>
<td>75</td>
</tr>
<tr>
<td>4.7</td>
<td>Distribution of the Total Path Length (TPL) between the Three Subject Groups</td>
<td>76</td>
</tr>
</tbody>
</table>
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 4.8</td>
<td>Distribution of the Average Distance from the Point of Interest (POI) or Centre of the Bench Model Between the Three Subject Groups</td>
<td>77</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>Distribution of the Average Distance from the Z-axis Between the Three Subject Groups</td>
<td>78</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>The Time Taken to Complete One-handed Knot Tying Skill Task of All the Trainees in the Study</td>
<td>93</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>The Total Path Length (TPL) of All the Trainees in the Study who Performed One-handed Knot Tying Skill</td>
<td>94</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>The Distance from the Z-axis of All the Trainees in the Study who Performed One-handed Knot Tying Skill</td>
<td>95</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>The Distance from the POI of All the Trainees in the Study who Performed One-handed Knot Tying Skill</td>
<td>96</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>The Time Taken by All the Trainees in the Study to Complete the Simple Interrupted Suturing Skill Task</td>
<td>97</td>
</tr>
<tr>
<td>Figure 5.6</td>
<td>The Total Path Length (TPL) of All the Trainees in the Study who Performed Simple Interrupted Suturing Skill</td>
<td>98</td>
</tr>
<tr>
<td>Figure 5.7</td>
<td>The Distance from POI of All the Trainees in the Study who Performed Simple Interrupted Suturing Skill</td>
<td>99</td>
</tr>
<tr>
<td>Figure 5.8</td>
<td>The Distance from Point of Interest (POI) of All the Trainees in the Study who Performed Simple Interrupted Suturing Skill</td>
<td>100</td>
</tr>
<tr>
<td>Figure 5.9</td>
<td>Difference in the Time Taken to Complete the One-handed Knot Tying Task at the Start of Training (Month 0 or Baseline) and at Five Months of Training</td>
<td>102</td>
</tr>
<tr>
<td>Figure 5.10</td>
<td>Difference in the Total Path Length (TPL) at the Start of Training (Month 0 or Baseline) and at Five Months of Training</td>
<td>103</td>
</tr>
<tr>
<td>Figure 5.11</td>
<td>Difference in the Distance from Z-axis at the Start of Training (Month 0 or Baseline) and at Five Months of Training</td>
<td>104</td>
</tr>
<tr>
<td>Figure 5.12</td>
<td>Difference in the Deviation Distance from the Point of Interest (POI) at the Beginning of Training (Month 0 or Baseline) and at Five Months of Training</td>
<td>105</td>
</tr>
</tbody>
</table>
List of Figures

Figure 5.13  Difference in the Time Taken to Complete the Task at the Beginning of Training (Month 0 or Baseline) and at Five Months of Training  107

Figure 5.14  Difference in the Total Path Length (TPL) to Complete the Task at the Beginning of Training (Month 0 or Baseline) and at Five Months of Training  108

Figure 5.15  Difference in the Deviation Distance from the Z-axis to Complete the Task at the Beginning of Training (Month 0 or Baseline) and at Five Months of Training  109

Figure 5.16  Difference in the Deviation Distance from the Point of Interest (POI) at the Beginning of Training (Month 0 or Baseline) and at Five Months of Training  110
### List of Tables

<table>
<thead>
<tr>
<th>Table 3.1</th>
<th>Participants Demographics for Hand Knot Tying and Suturing Skills</th>
<th>36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 3.2</td>
<td>The Item-test Correlation and $\alpha$ Values for Each Item in mGRS for One-handed Knot Tying skill</td>
<td>52</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>The Item-test Correlation and $\alpha$ Values for Each Item in mGRS for Simple Interrupted Suturing Skill</td>
<td>53</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>The Interpretations of Kappa, $\kappa$ Values according to Landis and Koch (1997)</td>
<td>54</td>
</tr>
<tr>
<td>Table 3.5</td>
<td>Inter-rater Agreement for Each Item in mGRS One-handed Knot Tying Skill</td>
<td>54</td>
</tr>
<tr>
<td>Table 3.6</td>
<td>Inter-rater Agreement for Each Item in Simple Interrupted Suturing Skill</td>
<td>55</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Participant Demographics for Hand Knot Tying and Suturing Skills</td>
<td>65</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Metrics of the Three Different Experience Groups</td>
<td>68</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>The Z Scores and $p$ Values from Mann Whitney U test for Each Metrics When Compared Between the Experts and Novices.</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Metrics of the Three Different Experience Groups</td>
<td>74</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>The Z Scores and $p$ Values from Mann Whitney U Test for Each Metrics When Compared Between the Experts and Novices</td>
<td>79</td>
</tr>
<tr>
<td>Table 4.6</td>
<td>Correlation Between the OSATS mGRS Scores and the Metrics from Patriot™ Motion Tracking System in One-handed Knot Tying Skill</td>
<td>81</td>
</tr>
<tr>
<td>Table 4.7</td>
<td>Correlation Between the OSATS mGRS Scores and the Metrics from Patriot™ Motion Tracking System in Simple Interrupted Suturing Skill</td>
<td>83</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>The Range, Mean and Standard Deviation of the Experts’ Scores for One-handed Knot Tying Skill and Suturing Skill</td>
<td>92</td>
</tr>
</tbody>
</table>
Summary

Background

Simulation technology is rapidly evolving and becoming the focus of attention in surgical training. However, the development of this technology in assessing open surgical skills is far behind when compared to minimally invasive surgery (MIS) training and assessment.

Aim

The overall objective of this thesis is to investigate the assessment tools that are readily available for assessing basic open surgical skills and develop an automated system that could be used in the simulation setting.

Materials & Methods

We evaluated the observer-dependant assessment tool in assessing basic surgical skills including one-handed knot tying and simple interrupted suturing skills. We established a hand tracking system that consists of an off-the-shelf motion analysis device and software created by our co-supervisor which generates numerical metrics. We investigated the validity of these metrics and we examined the potential application of this invaluable assessment tool.

Results

The results demonstrated that the observer-dependant assessment tool has poor inter-rater agreement despite its validity in assessing open surgical skills. The validity of the novel hand tracking system was established. We also demonstrated its application in assessing surgical trainees’ progress and also in proficiency-based assessment.

Conclusion

The hand motion tracking system is a valid assessment tool in evaluating performance in open surgical skills and has a potential significant role in the proficiency-based surgical training programme.
Acknowledgements

Supervisor
Professor Paul Neary

Co-Supervisor
Mr. Donncha Ryan

Additional Mentors
Professor Oscar Traynor
Mr. Declan Buckley

Statistician
Dr. Fiona Boland (The Biostatistical Consulting and Support Services in RCSI)

In Acknowledgement of their Generous Support
Mr. Kieran Tangney  Ms. Paula Mansell
Ms. Kate Gildea-Byrne  Ms. Eva Doherty
Ms. Emmeline Nugent  Ms. Emer Pyke
Ms. Jane Cunningham  Dr. Raj Sekhon

In Acknowledgement of their Time and Effort
All the Consultant Surgeons, Basic Surgical Trainees, Core Surgical Trainees and Medical Students of RCSI who participated in this study

Special Acknowledgement
Hj. Shaharan Tafri & Hjh. Zaharah Md Said
Zamshari, Shazleen, Shazrina & Shafeez
Adrian Foran
Chapter 1
Introduction
1.1 Background of Study

The nature of surgical training is consistently evolving in the past decade along with continuous changes in the healthcare system worldwide. The modern healthcare system has been pressurised by the current law that involves a restricted number of working hours. The legal working hours per week can be as low as 48 hours in the European countries (http://eur-lex.europa.eu/homepage.html, 2013) and 80 hours in North America (http://www.acgme.org). These working mandates are enshrined in legislation as it has been deemed necessary to guard against human errors. These may be related to stress and fatigue in a high-pressured working environment. The working time directive is a sensible precaution but it could cause a negative impact upon the surgical training of many aspiring surgeons.

In addition, there is also an increasing popularity in reporting medico-legal cases in the current media. This continues to lead on a working culture of zero complications from hospitals, medical insurers and members of the public. High profile medical reports such as the Kennedy Report (UK)(Kennedy, 2001) and the Institute of Medicine (US) report “To err is human” (Kohn et al., 1999) have highlighted surgical errors that turned the spotlight immediately on the adequacy of surgical training and, by extension, the quality of surgical trainees (Carroll et al., 2009).

Surgical educators are challenged to re-assess the training paradigm for educating surgeons to enable them to reach the highest standards in this new climate. Historically, surgical training was based on the apprenticeship model whereby the trainee surgeons were expected to take advantage of opportunities to learn and demonstrate their skills in the operating theatre. This was coined by William Halstead who exemplified the training approach as “see one, do one and teach one”(Cameron, 1997). This teaching method is largely relying upon variable cases that the trainees encounter during their daily work routine.

Such practice is less favorable in an age of restricted working hours and a zero complication ethos. The current climate in the healthcare system is limiting trainees the opportunities to gain sufficient experience and this could result in prolonged surgical training in order to achieve a high standard of qualified surgeons. This also affects the skills evaluation which is likely case-dependent.
and highly subjective. As a direct consequence of these challenges, interest in laboratories with formal curricula, specifically designed to teach surgical skills, has increased dramatically (Reznick and MacRae, 2006).

The use of surgical simulators and inanimate bench models for training and assessment has been a popular topic of discussion among the training bodies in Ireland and around the world to overcome this restriction. It offers many advantages to educators, especially at a time when trainee surgeons are finding it hard to obtain hands on experience and where skills assessments are subjective. Many innovative technologies are available to educators within the freedom of a more focused training set in a controlled and safe environment to the trainees. These learning environments provide the chance to practice the skills required of a surgeon in order to achieve a high standard of proficiency at their own pace. The greatest advantage of virtual reality medical simulation is the opportunity to try and fail without physical harm occurring to a patient (Satava, 2001). The surgical trainees are not in a pressurised environment and this helps them to develop their ability in various surgical procedures. Most importantly, the patients’ safety is not being jeopardized rather their safety is improved and protected from any incompetency or misconduct from surgeons in training.

The integration of simulation technologies into training programmes would therefore seem the logical step for the design and implementation of any modern surgical training curriculum. Alongside the continued development in surgical training is the ability to assess the performance of the surgical trainees in technical skills adequately and objectively. Assessment can be defined as making a judgement against a predefined reference (Beard, 2008). In surgical assessment, candidates are assessed in order to determine how well they are able to meet certain technical criteria and their level of proficiency (Ahlberg et al., 2007). Historically, under previous methodologies for teaching and the acquiring of surgical skills, this assessment has been very largely subjective. It has been based upon the opinions and views of experienced surgeons or surgical educators. This has long been recognised as a problem, especially in the modern context, where zero complications are expected and demanded. The development of technologies such as simulators can possibly offer some
solution and help to ensure that the evaluation of any candidate is as thorough, fair, transparent and objective as possible.

The objective characterisation of technical skills can be very difficult, even for those who are very experienced in their examination. This is also the case with the skills required of a modern surgeon. There is a wide array of skills required of a surgeon that need to be evaluated. This assessment of technical performance ranges from basic surgical skills such as knot tying and suturing, basic laparoscopic skills and endoscopy, to a wide spectrum of evaluations that include performing complex procedures such as laparoscopic cholecystectomy, vessel anastomosis and tendon repair.

This thesis will address the issue of change in the evaluation of surgical training. There is a need to improve the appraisal of surgical ability in order to ensure that the highest standards are maintained in the training of surgeons. The evaluation of the surgical ability of trainees is also essential, in order to enable trainees to receive the best training. This may be achieved by improving the quality of the feedback that they receive based upon assessment of their performance of surgical tasks. This research will examine the potential benefits of non-observational tools and simulators in the assessment of surgeons as they undergo their training. This will involve an assessment of OSATS observational tools and examine their reliability and objectivity when it comes to evaluating the skills of trainee surgeons and compare these with non-observational tools.

Technology is offering the prospect of the objective assessment of surgical training. Its use may permit moving away from observational subtype tools. Increasingly, there are new technologies that allow for the objective assessments of surgical trainees. These do not require observers and may ultimately be more objective, transparent, reliable and reproducible. The rationale for this research is to determine if simulators can provide better evaluations of surgical trainees. This is important, for the training of surgeons as it can highlight their strengths and weakness. The feedback provided by the improved evaluations of their skills will allow those training in surgery to have better learning outcomes. This is because these non-observational assessment
tools will provide more objective and consistent findings with regard to trainees’ skills level.

1.2 Context and surgical training

Until the 19th century, the most common method for training surgeons was through apprenticeships. This involved an aspiring doctor working with a well-established surgeon and receiving on-the-job training. The apprentice would learn surgery by direct observation. The training was not systematic and largely, it became opportunistic or fortuitous in the real hospital setting. In the 19th century, all medical science advanced, surgeons became increasingly professional. All surgeons are required to be fully licensed and registered (The Medical Act 1866). They were obliged to undertake a minimum level of training and undertake examinations in order to be registered as a surgeon. By the start of the 20th century, there was an internationally recognized teaching model for surgery which is known as the Halstedian method (Cameron, 1997). As mentioned previously, this method involves surgeons learning a science-based curriculum during their day-to-day work which included bedside ward rounds and supervised surgery.

The development of minimally invasive surgery over the last few decades has been revolutionary and it has become a feature of every area of surgery. Technological advances, including laparoscopic instruments, to stapling devices, to endoscopic technology, enhanced the capabilities of surgeons but also created challenges for surgical educators. Today, surgical trainees need to learn a significant amount of technical information and receive extensive training in the safe and efficient use of various new instruments and technologies. This means that surgical education is even more demanding and that there is an imperative to ensure that it is of the highest standard. Objective and consistent evaluation of surgical skills are an important part in any effort to improve the general level of training available to those who intend to pursue a career in surgery. This is not only in minimally invasive surgery but more importantly in the basic open surgical skills such as suturing and knot tying. However, there is a gap in the literature in regards to the development of simulation in assessing open skills.
Therefore, the main focus of this thesis is to assess where we currently stand in relation to the use of simulation in open surgical skills assessment within current training curricula. It is centred upon the use of automated skills assessment in evaluating open surgical skills within surgical curricula in order to establish proficiency profiles of surgical trainees in a safe laboratory environment.

1.3 Technology and Surgical Training

It is important to assess trainees on their skill levels and competency. This is in the interest of both the candidate and the patient. This ensures that they can meet the high demands expected of them and that they are able to minimize or eradicate the risk of a patient’s complication as a result of a surgical procedure.

The potential use of simulator technologies in particular, allows the trainers to give consistent and constructive feedback based on their performances. They also can be used for the examinations of surgical skills and functional capability. Despite its importance to surgeons, technical proficiency historically has been poorly evaluated (Ahlberg et al., 2007). There has been a need for a more objective assessment tool available to surgical educators. A good assessment tool must possess reliability, validity, educational impact, acceptability and feasibility (Schuwirth and van der Vleuten, 2004).

A virtual reality (VR) surgical simulator is a type of computerized technology that is developed to imitate, in as realistic manner as possible, surgical procedures for the purpose of training. Since the advent of minimally-invasive and laparoscopic surgery, many surgical simulators have emerged to train surgical trainees. The first simulators dated from the 1980s with the utilization of video games technology, especially the graphics. Surgical simulators are used to train doctors in surgical procedures. These ideally are without the use of animals or cadavers. It allows them to acquire a predefined level of skills and demonstrate proficiency before carrying out a procedure with real patients.

The VR simulators are created to provide objective measurement of a specific surgical skill. In general, the simulator uses a computer screen displaying a two-dimensional graphic of the internal organs being operated upon. The surgical instruments used are fitted with sensors that allow the cameras to track their movement. From this, the simulators are able to quantify and converts the
movement to meaningful metrics such as path length, smoothness and economy of movement. Some examples of validated VR simulators available in laparoscopy are MIST VR, LapSim, LapMentor and Xitact LS500 (Schijven and Jakimowicz, 2003). These simulators could assess various laparoscopic skills from basic tasks such as object positioning and manipulation to more complex tasks such as laparoscopic cholecystectomy. The main criticism of VR simulators however is that they lack convincing elements of real life representation such as delayed gravity effect and no haptic feedback, as found in LapSim (Munz et al., 2007).

The simulation technology continues to develop a hybrid simulator that provides a true tactile feedback which is lacking in VR simulators. The ProMIS™ (Haptica, Dublin) used 100% VR for certain tasks and augmented reality that overlays graphics onto a task performed on a physical exercise (Buckley CE, 2012). Apart from providing movement-based metrics, it is also capable of identifying errors such as number of dropped pegs in peg transfer task and accuracy in precision-cutting task.

The use of motion sensors to track physical movements has finally allowed simulators to simulate the real situation in a surgical environment. The advancement in simulator technology provides a relatively realistic simulation of a surgical procedure. They are in particular suited to monitoring eye-hand coordination abilities of a trainee. Eye–hand coordination is improved because the simulation can give both visual feedback, by way of a screen, as well as tactile feedback that simulate the manipulation of organs and tissue.

These simulators are mainly used for training purposes, especially in the training of surgeons in laparoscopic procedures. It is only in recent years that they have been incorporated in the skills assessment and trainees’ selection process. This is very different to other industries such as the aviation industry that are pioneers of simulation-based training. Out of 52 articles in the literature, 56% of the studies employed simulator-generated objective metrics in the laparoscopic skills assessment, either exclusively or combined with other assessment tools (Shaharan and Neary, 2014).
Apart from laparoscopic skills, surgical trainees, especially those training in general surgery, are required to obtain endoscopic skills too. Training in endoscopy in a virtual environment is thought to be a good alternative to classical bedside teaching, but without the adverse effects, such as patient discomfort, risk of perforation, and longer procedure time (Grantcharov et al., 2005). One of the endoscopic simulators commonly used in surgical training centres is the GI Mentor (Symbionix, Israel). It provides detailed evaluation on the trainees’ performance including time taken to complete the task, percentage of mucosa visualized and percentage of time spent without clear vision (red-out) (Moorthy et al., 2004).

The advancements in simulation technology have allowed The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) to develop training models for both laparoscopic and endoscopic skills. The Fundamentals of Laparoscopic surgery (FLS) is a validated program for the teaching and evaluation of the basic knowledge and the skills required to perform laparoscopic surgery (Vassiliou et al., 2010). It is a comprehensive training model which includes continuous didactic teachings of the trainees and structured practical skills using the laparoscopic simulators. Currently, it is a requirement to complete the FLS program for American Board of Surgery for certification in general surgery (Rooney et al., 2012). The development of simulation-based training continues to embrace the training and assessment of endoscopic skills. The Fundamental of Endoscopic Surgery™ (FES™) was developed after the FLS which serves to teach and assess the fundamental knowledge and skills required to practice flexible endoscopy of the gastrointestinal tract (Poulose et al., 2013).

In general the progression of simulator development has tended to target minimally invasive surgery (MIS) (Neary et al., 2008). The development of simulators to train basic surgical skills in open surgery remains a challenge. Competency in basic surgical skills such as knot tying and suturing is vital in surgery to ensure zero complication in patients’ care. The surgeon’s ability to tie knots securely is of paramount importance as loosening of surgical knots during or after tying can compromise the outcome of a surgical procedure (Alzacko and Majid, 2007). The acquisition of basic open skills forms a strong foundation
in the trainees' ability to progress into performing more complex procedures such as laparotomy, tendon repair and vessel anastomosis.

Nevertheless, the training and assessment of the open surgical skills are largely depending on inanimate bench models. Some examples of commonly used bench models are skin pads and saphenofemoral junction model from Limbs and Things™ (Bristol, United Kingdom) and laparotomy model from Simulab Corporation (Seattle, WA). The challenges in developing automated simulators for open surgical skills include what parameters should be recorded, how to collect the data and interpret it to a meaningful metrics. Therefore, the technology in measuring the competency in open surgical skills is slowly evolving.

1.4 Observer Dependant Assessment Tools

Observer dependant assessment tools remain the gold standard for assessing the fundamental open technical skills such as suturing and knot tying. The classic observational assessment tool for open surgical skills is the Objective Structured Assessment of Technical Skills (OSATS). It was coined by Professor Reznick and his research team in Canada (Martin et al., 1997). It is based on two sets of checklists. The first part is a checklist that is specific to the operation or procedure that is being assessed. It consists of a list with specific important steps in the procedure and each step is scored whether the task was done correctly or not (Figure 1.1). The second part is another set of checklist that examines the global performance of the subject rather than specific steps or tasks. Hence, it is called Global Rating Scale (GRS). Typically, the performance is assessed by looking at the essential elements outlined in (Figure 1.2).

The OSATS is consistent with the format of the typical Objective Structured Clinical Examination (OSCE) in which examinees perform a series of clinical tasks at each of several time-limited stations (Paisley et al., 2001). Typically, a trainee will perform a specific procedure such as excision of sebaceous cyst using an inanimate model and an observer who has extensive experience in the field such as a consultant or a senior registrar will use the OSATS checklists to score the trainee's skill. This can be done either face-to-face or video recording (Jensen et al., 2008).
### Appendix 1

**Checklist For Skin Suturing**

Name: [Student Name]  
Years of Training: [Years]

**Instructions to candidates**

<table>
<thead>
<tr>
<th>Item</th>
<th>Done correctly</th>
<th>Not done correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Selects appropriate suture, needle holder and forceps.</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2. Needle loaded ½ to 2/3 from tip.</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>3. Bite distance from the skin edge-5mm.</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4. Angle at which bite taken - 90°</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>5. Single attempt while taking bites in the skin</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>6. Movement occurs at wrist</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7. Forceps used to hold skin or s/c tissues (minimum use)</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8. Whether takes bites from both skin edges in one go/separately</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9. Equal bites on both sides</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10. Whether needle touched with hand</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11. Number of knots taken</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>12. Knot is square or not</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>13. Knot is too tight or too loose</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>14. Suture breaks or not</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>15. Knot is on the incision line or on one side</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16. Distance of cutting the suture from the knot</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>17. Suture board moves or not</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>18. Skin edges are everted or inverted</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19. Intersutural distance - 0.5 to 1cm.</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Maximum Total Score**

<table>
<thead>
<tr>
<th>Total Score</th>
<th>Examiner</th>
</tr>
</thead>
<tbody>
<tr>
<td>(19)</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.1 A sample of task-specific checklist from OSATS (Shindholimath VV, 2003).*
There is also a different type of observational tool which was developed by a research team in Quebec, Canada and it is called Global Operative Assessment of Laparoscopic Skills (GOALS) (Vassiliou et al., 2005). Similar to GRS from OSATS, GOALS consists of 5 items assessing the global performance of the trainee (Figure 1.3). It is scored by using the 5-Likert scale with anchors at 1, 3, and 5 where “1” represents the lowest level of performance and “5” is considered ideal performance.
# Global rating scale component of the intraoperative assessment tool*

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

**Bimanual dexterity**
- 1. Uses only one hand, ignores nondominant hand, poor coordination between hands
- 2.
- 3. Uses both hands, but does not optimize interaction between hands
- 4.
- 5. Expertly uses both hands in a complimentary manner to provide optimal exposure

**Efficiency**
- 1. Uncertain, inefficient efforts; many tentative movements; constantly changing focus or persisting without progress
- 2.
- 3. Slow, but planned movements are reasonably organized
- 4.
- 5. Confident, efficient and safe conduct, maintains focus on task until it is better performed by way of an alternative approach

**Tissue handling**
- 1. Rough movements, tears tissue, injures adjacent structures, poor grasper control, grasper frequently slips
- 2.
- 3. Handles tissues reasonably well, minor trauma to adjacent tissue (ie, occasional unnecessary bleeding or slipping of the grasper)
- 4.
- 5. Handles tissues well, applies appropriate traction, negligible injury to adjacent structures

**Autonomy**
- 1. Unable to complete entire task, even with verbal guidance
- 2.
- 3. Able to complete task safely with moderate guidance
- 4.
- 5. Able to complete task independently without prompting

* The descriptors shown are the “anchor” descriptors for scores 1, 3, and 5

---

**Figure 1.3: Global Assessment of Laparoscopic Surgery (GOALS) scoring system** (Vassiliou et al., 2005).
In our institution, we use OSATS checklists to assess open surgical skills in all our trainees at different level. We also utilise OSATS as part of the selection process into the higher surgical training (HST) programme. All candidates shortlisted for the HST programme in general surgery, cardiothoracic and plastic surgery is required to go through surgical skills assessments prior to their interviews (Shaharan and Neary, 2014).

This type of assessment tool requires expert surgeons to utilise the tool when assessing surgical skills. The tools convert subjective assessments to a set of an objective scoring scheme. Therefore, it is open to critique such as human bias or favouritism during the assessment. In order to minimize such bias, it is required to have at least two expert surgeons assessing the trainees independently. The degree of discrepancy between the assessors should be kept to low level. This can be analysed by calculating the inter-rater (IR) reliability. If the IR reliability value is 0.8, this means that 80 percent of the scores among assessors were in agreement and 20 percent were in disagreement. A high value of IR reliability indicates that the scores are homogenous and the assessment tool is both robust and of value (Shaharan and Neary, 2014). Therefore, in order for the assessment to be valid and useful, it is recommended that multiple assessors are required to utilise this method in assessing surgical skills. This is obviously the main disadvantage of the observational tools as it is labour intensive. We would argue that the use of objective assessment using simulators solely would be a stronger approach to assessment the vital skills in surgery.

In this thesis, we therefore explained the use of OSATS and compared it to an objective assessment using validated simulation-based technology.

1.5 Objectives

1.5.1 Hypothesis Underlying the Objectives

The overall objective of this thesis is to investigate the assessment tools that are readily available for assessing basic open surgical skills. Traditionally, observer-dependant assessment tool such as OSATS has been largely used to assess surgical trainees throughout their training programme. As we are moving
closer towards era of simulation training, there is paucity in the development of technology in evaluating open surgical skills which is fundamental especially in the early stage of training. The idea is to develop a potential assessment tool which is automated and does not require any assessors or observers to evaluate the trainees’ skill performance. This would be more pertinent and valuable as we are moving towards the direction of proficiency-based training curriculum.

1.5.2 Detailed Objectives

Objective 1: To evaluate the validity and reliability of the Global Rating Scale (GRS) from OSATS

The GRS consists of components that assess the general performance during skill assessment. This assessment tool requires multiple expert observers to evaluate trainees’ skill in open surgical technique. We aim to evaluate the validity and reliability of this tool in assessing basic surgical skill,

Objective 2: To explore and investigate the validity of a motion tracking system for assessing basic technical skills.

An automated assessment tool would be able to generate numerical metrics when assessing the hand motion. This tool would not require any expert observers to assess the technical skill performance. We aim to evaluate the validity of metrics produced by a motion tracking system in order to analyse trainees’ performance in open surgical skills. This potentially could eliminate the need to recruit assessors during surgical skill assessment.

Objective 3: To explore the application of a motion tracking system in surgical training curriculum.

The application of a motion tracking system in surgical training is explored as a follow on from validating the system in assessing trainees’ performance in basic open surgical skills. As the system produces numerical metrics, it would be feasible to use it in the surgical training curriculum. We aim to determine the use of the motion tracking system in monitoring surgical trainees’ progress in the surgical training process.
Chapter 2

Materials and Methods
2.1 Ethical Consideration

Ethical approval for this study was granted by the Research Ethics Committee of RCSI. Each volunteer was provided with a subject information sheet containing the details of the study and what to expect during their involvement. All volunteers were required to give written informed consent according to an agreed protocol by the ethics committee. All data collected will be stored in an anonymous format by using numerical codes. The volunteers were assured that all data was treated as confidential and was not to be shared with any third party such as training supervisors or hospitals.

2.2 Recruitment of Participants

An invitation email was sent to all RCSI medical students in pre-clinical years (Year 1-3), basic surgical trainees (Year 1 and Year 2) and consultant surgeons who were involved in teaching surgical trainees in the surgical skills lab. Posters were displayed in the surgical skills lab and lecture theatres in RCSI. Additional announcements were made before and after operative skills classes and lectures.

It was made clear that participation was voluntary. Random selection of participants was performed from the responses received. All selected participants were contacted via email to arrange a time that most suited them to attend the assessment. The assessment was performed in the National Surgical Training Centre (NSTC), Royal College of Surgeons in Ireland, 121 St. Stephen's Green, Dublin 2.

2.3 Participant Demographics

A total of 27 medical students, 24 basic surgical trainees and 10 surgical experts were recruited. Detailed demographics for the participants are outlined in each chapter.
Medical students who were in pre-clinical years 1-3 in RCSI were included in this study. Students with previous experience in surgical skills such as during a summer elective were excluded. This was to ensure that the subject group was a complete novice group.

The inclusion criterion for surgical trainees was trainees in Basic Surgical Training (BST) programme Year 1 and Year 2 of their postgraduate training. They also must have completed a basic surgical skills course.

For the expert group, we invited consultant surgeons who had experience as a consultant surgeon for greater than 5 years. These surgeons were from different specialties including general surgery, plastic and reconstructive surgery and paediatric surgery.

2.4 Basic Surgical Skills Assessment

Basic surgical skills were assessed using standardised bench models. The bench models used in the assessment were from Limbs & Things™ (Bristol, UK) which include the knot tying trainer jig, skin pads, skin pad jig and cyst pads. The skills examined were as follows:

Task 1: One handed knot tying skill

Task 2: Simple interrupted suturing technique

Verbal instructions were given prior to the assessment and written instructions (appendix I) were provided to subjects for each task. Subjects were allowed to refer to the instruction sheet at any time during the assessment. Subjects in the novice group were shown videos on each skill so as to familiarise themselves with the tasks. Each subject group was given equal opportunity to practise the skills prior to the assessment. Subjects were required to repeat each task 3 times.
Task 1 Knot tying skill: One-handed technique
The subjects were required to perform surgical knots using the one-handed technique. They were given a 2/0 Mersilk (Ethicon) tie with a standard length of 46 cm each. The subjects were instructed to tie the surgical knots on the knot tying trainer jig. The knot tying trainer consists of two red rubber tubes marked with 5mm white tape which was secured at the centre of the tubes (Figure 2.4). Subjects were instructed to tie surgical knots within the white marking tape with 6 throws.

Task 2 Suturing skill: Simple interrupted suturing technique
The subjects performed simple interrupted sutures. Each subject was given a 3/0 Mersilk suture (Ethicon) with standard length of 30 cm. The subjects performed these tasks on skin pads (Limbs & Things, Bristol, UK). Skin pads were prepared by making 1 cm longitudinal incisions to simulate the wound edge and two round marks were drawn 0.5 cm from the edges with permanent marker (Figure 2.5). The subjects were instructed to pass the suture through the confines of the marked area on the skin pad and then perform an instrument tie with 3 throws. It was made clear that the skin edges should be approximated and there should not be any gap between the knot and the skin pads.

2.5 Assessment Tools
2.5.1 Observational Assessment Tool
The Global Rating Scale (GRS) from OSATS (Martin et al., 1997) was used as an observational assessment tool. It is a generic tool used in assessing many surgical procedures. It consists of 7 different components and each of them is rated according to a Likert scale, 1 being poor performance and 5 being excellent performance (Figure 2.1). Some of the components in the GRS were thought to be unsuitable for the skills assessed in this study. Therefore, the GRS was modified by selecting appropriate components according to the specific skill and this was labelled as modified-GRS (mGRS).
For knot tying skill, only three components were included in mGRS: “respect for tissue”, “time and motion” and “flow of operation” (Figure 2.2). In the assessment of simple interrupted suturing skill, all components were included except “use of assistant” because it was deemed to be irrelevant.

Videos were recorded when the trainees performed the surgical skills. The camera was focused on the bench models and at the subjects’ hands. The microphone was switched off to eliminate any potential identification of the subjects. Each video was anonymized by a random code generator.

Two assessors were selected to assess each video using the mGRS. The assessors were expert surgeons with greater than 10 years of consultant experience and are involved in teaching and educating surgical trainees in Ireland. The videos were played in random sequence so that the observers could not detect any pattern according to the subject groups. Figure 2.7 outlines the experiment process.
Figure 2.1 The original Global Rating Scale (GRS) in the Objective Structured Assessment of Technical Skills (OSATS).

The GRS consists of seven items in rating the operative performance according to the Likert scale. (Reznick et al., 1997) The modified-GRS (mGRS) for knot tying skill assessment consists of 3 items which are “respect for tissue”, “time and motion” and “flow of operation” and for simple interrupted suturing skill, all items were included except “use of assistant”. 
Figure 2.2 The modified Global Rating Scale (mGRS) for knot tying skill assessment.

The mGRS consists of three relevant items in rating the operative performance according to the Likert scale.

2.5.2 Non-observational Assessment Tool

The Patriot™ (Polhemus Inc., Colchester, VT) motion tracking system was used in this study as a non-observational tool. It is a commercially available device using electromagnetic technology that is widely used in areas such as training spray painters and ultrasound simulation (http://polhemus.com/motion-tracking/all-trackers/patriot).

It consists of a sensor, an electromagnetic source and a System Electronics Unit (SEU). The SEU contains the hardware and software necessary to generate and sense the magnetic fields, compute position and orientation, and interface with the host computer via an RS 232 or USB interface (http://polhemus.com/_assets/img/PATRIOT_Brochure.pdf). The system tracks 6 degrees of freedom (6DOF) measurements of the sensor’s movement and
collects the raw data into the PiMgr software (Figure 2.3). It has an update rate of 60 Hz per sensor for an accurate measurement.

For this study, the sensor was attached to the subject’s right index finger with a Velcro® strap (Figure 2.6). The position and orientation of the sensor was recorded at the starting point of each skill assessment until the last step of the examined task. Bespoke computer software was created by our co-supervisor, Mr. Donncha Ryan to translate the raw data into simple metrics. The metrics generated from the in-house software were time, total path length (TPL), deviation distance of the sensor from x-, y-, z-axis and from the point of interest (POI). The POI was set up as the central point of the bench model. This was done by capturing the position of the sensor when it was static in the middle of the bench models. This fixed position was entered into the in-house programme prior to analysis. Figure 2.8 outlines the experiment process.
Figure 2.3 A screenshot of the PiMgr software.

The airplane image (lower part of the screenshot) indicates the sensor which will move to the position and orientation of the right index finger. The retrieved position and orientation are displayed as numbers in six columns (upper part of screenshot), from left to right, positions in X-, Y-, Z-axis and orientation in yaw, pitch and roll.
Figure 2.4 The set-up for knot tying skill assessment with the Patriot™ motion tracking system.

The Patriot™ system consists of the System Electronic Unit or SEU (top left corner), a source (top middle) and a sensor (right top corner of the knot tying jig).
Figure 2.5 The set-up for simple interrupted suturing assessment with the Patriot™ motion tracking system.

Incisions were made on the skin pad and marked with permanent marker. A needle holder, suture, toothed forcep and suture scissors were provided.
Figure 2.6 The Patriot™ sensor attached to a trainee’s index finger while performing simple interrupted suturing task.
The instruction sheet was available on the top right of the photo.
2.6 Application of Metrics from Patriot™ Motion Tracking Device

Once validated, these metrics were applied in a proficiency-based assessment of trainees' performance, as outlined in Figure 2.9. It was based on a proficiency goal determined by a group of experts. The proficiency level for each task was calculated by adding the mean scores of the experts and the standard deviation (Proficiency score = Mean + 1SD). The trainees’ scores were mapped out for each metric against the proficiency level. This identified any trainee who was unsuccessful in reaching the desired goal and required remedial intervention in order to achieve competence in the skill.

We also used this method to assess the progress of newly appointed basic surgical trainees. Their basic skills were examined during their mandatory bootcamp and first clinical rotation. The assessment was repeated after 6 months following their clinical job appointment. Their performance was then analysed according to the proficiency level determined as described above.

2.7 Statistical Analysis

All data was analysed using Stata software (StataCorp LP, Texas, USA). All numerical data was tested for its normality in distribution with Shapiro-Wilk W test. Parametric tests such as paired t-test were used in normally distributed data and non-parametric tests such as Kruskal-Wallis test and the Mann Whitney U test were used for data which was not normally distributed.

The internal consistency of the mGRS was measured using Cronbach’s alpha. Inter-rater reliability between the two examiners for mGRS scores was calculated using kappa (κ) test.

For construct validity, Kruskal-Wallis test were used to determine any significant difference between the three study groups for each metric. When there was a significant difference, the performance between experts and novices was analysed with Mann Whitney U-test to further assess the validity of the metrics. Concurrent validity was analysed by correlating the metrics for each group using Spearman correlation.
The metrics and mGRS scores had normal distribution. The paired t-test was used to analyse the metrics, including the GRS, before and 6 months after training. We used the Pearson product-moment correlation coefficient to compare the correlation between the metrics and the OSATS score.
Figure 2.7 Flow diagram of the experiment process in Chapter 3 - Validity and Reliability of Global Rating Scale in Open Surgical Skills.
Figure 2.8 Flow diagram of the experiment process in Chapter 4 - Validity of a Motion Tracking System in Basic Surgical Skills.
Figure 2.9 Flow diagram of the experiment process in Chapter 5 - The Application of Patriot Motion Tracking System in a Proficiency-based Assessment of Basic Surgical Skills.
Chapter 3
Validity and Reliability of Global Rating Scale (GRS) in Open Surgical Skills
3.1 Introduction

Open surgical skills such as knot tying and suturing remain as fundamental skills in surgery. Trainees are trained in these skills at the earliest stage possible. Competences in these skills are repeatedly evaluated either formally or informally during training to ensure trainees are safe to operate in a real life setting. The main learning points for trainees to achieve competence include economy of movement, smoothness, efficient movement and effective tissue handling which includes minimising trauma to soft tissues. These key competencies ensure that trainees are comfortable in performing basic skills with minimal complications for the patient.

In order to properly monitor the education of trainee surgeons it was necessary to find a reliable method to evaluate their skills. Traditionally the evaluation of the trainee surgeons skills were judged by their superiors. This is highly subjective, who usually provided the trainee with feedback. This has since been replaced by several more objective assessments of trainee surgeons’ surgical skill levels.

The commonest assessment tool used is the OSATS Global Rating Scale (OSATS GRS). The OSATS GRS is a classic observational assessment tool developed by Professor Reznick and his fellow colleagues (Reznick et al., 1997). It is a generic assessment scoring scale that is widely used in evaluating trainees' performance in many surgical procedures (Shaharan and Neary, 2014). This includes evaluating skills in open surgical procedures such as excision of skin lesions and wound closures (Jensen et al., 2008, Chipman and Schmitz, 2009) and laparoscopic procedures such as laparoscopy cholecystectomy (Arora et al., 2011, Lucas et al., 2008, Palter et al., 2013) and laparoscopy jejunostomy (Varas et al., 2012). It involves recruiting assessors who are experts in a particular skill that is being evaluated. They will observe the trainee and judge their skills based on their observation. The performance of each trainee is assessed, either live during the assessment or by video recording. The raters are expected to give their opinion instantaneously. This subjective scoring usually does not involve inter-rater reliability test, most likely due to the manpower required in subjective assessment.
The scores of the OSATS GRS are used to provide information on the progress of a trainee, their skills sets and competency. It is necessary to determine the trainees’ level of surgical skill competency given the nature of the training and its implications for the health of patients. Therefore, it is crucial that any evaluation system to assess their competency is valid and reliable.

This chapter will test the validity and reliability of the OSATS GRS in assessing basic surgical skills as a good assessment tool. It will seek to test if the tool is suitable for the measurement of these fundamental skills. Every evaluative tool needs to provide reliable information on what it measures or examines and that the conclusions drawn from the tool are dependable.

Reliability is the concept that a measurement tool can achieve reproducible results among users and at different points in time (Higgins and Straub, 2006). This concept is important in assessing the critical skills of surgical trainees. Surgical skill assessment tools need to be consistent and replicable. If a measurement is unreliable it could mean that an incompetent individual will be incorrectly judged to be proficient in surgical skills and this could have dire consequences for patients and hospitals (Reznick and MacRae, 2006).

To determine if the rating system is reliable and produces consistent and replicable data with relation to the level of competency of trainee surgeons, two key criteria will be discussed.

1. **Internal consistency**
   Internal consistency measures how well the items on a test measure the same construct or idea (Williams, 2003). The aim of the GRS is to assess general surgical principles (Cao et al., 1999). The assessment scale provides an evaluation of the general performance of an individual trainee. It typically consists of 7 evaluation items including respect for tissue, instrument handling and flow of operation. Each item is scored using 5-Likert type scale, with 1 being poor performance and 5 being excellent performance. These items should correlate with each other to ensure that they measure the same general idea which is the competency of the trainee surgeons in basic surgical skills.

2. **Inter-rater reliability or agreement**
The GRS is scored by the assessor according to the degree of knowledge, skills and dexterity of the trainees. The assessors score the trainee based upon their own training and experience, which is highly subjective. As an observational tool, it was recommended that at least two assessors or observers are required at a given time to assess one subject or trainee (Acock, 2012). The reliability of the GRS will be assessed if it produces consistent measurements of skills level among the assessors. The inter-rater reliability examines if there are any significant differences between the two assessors. The GRS will be deemed reliable if the assessors are generally in agreement on their scoring of the skills competency level. The reliability of the GRS will need to show that the assessors are generally consistent in their scoring but not necessarily identical. The inter-rater reliability would prove that in general the scores must be broadly similar and uniform. This is to prevent any erratic observer which results in unreliable score and reduce the chance that an assessor’s subjective opinions and views could possibly distort the findings.

Another important element in assessment tool is construct validity. One inference of construct validity is the extent to which a test discriminates between various level of expertise (Moorthy et al., 2003). A validated assessment rating scale should be able to differentiate level of surgical skills according to the level of competency.

There are both advantages and disadvantages of OSATS. Among the advantages of the instrument is that it can be applied to a wide range of surgical procedures, including surgical skills. It is a relatively quick and easy tool. The data collected by this method are also easy to collect and to enter into a database. It also can provide feedback on the performance. On the other hand, the assessors need to be trained on how to use it and how to score the trainees. There may be different scores provided by the evaluators due to unintentional bias, particularly if the raters have any information about those being assessed. In order to overcome this, it is recommended to have at least two independent assessors. Therefore, it is time consuming and labour intensive.
Research conducted by Martin et al show that the scoring system is quite reliable and that its internal consistency was in the range of (alpha: 0.61-0.74) and this was generally a moderate to high score (Martin et al., 1997). This preliminary study suggests that the OSATS can reliably and validly be used to assess surgical skills. It has limitations. There is an agreement that it should not be used as a stand-alone assessment tool but it needs to be complemented by other metrics. This will then ensure that a comprehensive assessment of a trainee’s skills can be found, that is highly reliable.

In this chapter, our aim was to ascertain the reliability of the GRS in the assessment of basic open surgical skills. The basic skills that were included in this study were one-handed knot tying and suturing task of simple interrupted skill. The chapter is not concerned with the findings with regard to these two skills but as to how reliable the GRS system is in assessing these two fundamental skills.

3.2 Objectives

1- To assess the construct validity of the OSATS GRS that has been modified for two basic skill set (hand knot tying and simple interrupted suturing).
2- To analyse the homogeneity and consistency of the components in the OSATS GRS.
3- To determine the inter-rater reliability of the OSATS GRS for both skill set in the most ideal setup.

3.3 Materials and Methods

3.3.1 Recruitment of Participants

We recruited consultant surgeons as experts, basic surgical trainees year 1 and year 2 and medical students as novices, as explained in Chapter 2. This provided the study with a range of competency which was evaluated by the modified Global Rating Scale (mGRS). There were differences expected in the skill levels of the various participants. The study was particularly focused on the
ability of the mGRS to differentiate the various skills levels among the participants, such as that between an expert and a trainee.

3.3.2 Participant Demographics

A total of 28 trainees and 25 novices were included in hand knot tying skill and 16 trainees and 27 novices were included in suturing skill (table 3.1). Nine experts were invited to participate in the study but four were excluded from knot tying skill because they formed the loops with their right hand.
Table 3.1 Participant demographics for hand knot tying and suturing skills.

<table>
<thead>
<tr>
<th></th>
<th>Expert</th>
<th>Trainee</th>
<th>Novice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hand Knot Tying Skill</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Dominant Hand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>4</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>Left</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>

|                          |        |         |        |
| **Suturing Skill**       |        |         |        |
| Gender                   |        |         |        |
| Male                     | 8      | 10      | 18     |
| Female                   | 1      | 6       | 9      |
| Dominant Hand            |        |         |        |
| Right                    | 9      | 15      | 24     |
| Left                     | 0      | 1       | 3      |
3.3.3 Skills Assessment Tool

A modified Global Rating Scale (mGRS) was used to assess the surgical skills, as described in Chapter 2. This involved selecting a number of items assessed which were concentrated on two main basic surgical skills: one-handed knot tying and simple interrupted suturing skills. These are two most common skills which would allow us to assess the mGRS for its reliability and validity. It is acknowledged that one of the strengths of the GRS is that it is extremely flexible when evaluating the surgical levels of trainees and others. For knot tying skill, three items only were included: “respect for tissue”, “time and motion” and “flow of operation”. In the assessment of simple interrupted suturing skill, item “use of assistant” was excluded as it was deemed to be irrelevant.

3.3.4 Surgical Skills Assessment

The skills that were assessed in this chapter were one-handed knot tying skill and simple interrupted suturing skill. Each subject was required to perform these skills three times in order to achieve an average of the three trials. This was to provide a better understanding of the trainees’ skills in general and therefore preventing any atypical performance of the skills which could distort the assessors’ evaluation. This was necessary in order to ensure the reliability of the findings. Anonymous video recordings were taken during the assessment, as described in Chapter 2. This enabled the assessors to evaluate the skills of the trainees in real time and yet they did not have to be present when they were actually performing the tasks. Two experts (PN and DB) assessed the video recordings independently. This is the minimum number of assessors recommended to evaluate the GRS assessment tool.
3.3.5 Statistical Analysis

Statistical analysis was performed using Stata 12 (StataCorp LP, Texas, USA) on the assessors’ evaluation. This analysed the scoring on the Likert scale of the two blinded assessors. The internal consistency of the mGRS was measured using Cronbach’s alpha. Inter-rater reliability between the two examiners for mGRS scores was calculated using kappa (κ) test.

3.4 Results

Total of 171 video recordings of one-handed knot tying and 156 video recordings of simple interrupted suturing skill were viewed and assessed by the two blinded raters (PN, DB) using the mGRS score. Three knot tying videos were excluded due to poor quality. The video recordings provided the evidence that was the sole basis for the evaluation. The possibility that this may have impacted upon the reliability of the rating system will also be discussed.

3.4.1 Construct Validity of mGRS for One-handed Knot Tying Skill

Figures 3.1-3.3 showed the distribution of the scores for each item in mGRS. There was a significant difference in all of the three components of mGRS between experts, trainees and novices, (p<0.05). The experts consistently scored the highest mark, followed by the trainees and novices. This was expected and conformed to the findings of the literature. This was consistent over the three markedly different components of the ratings. The three components were important skills which required different abilities and the fact that the findings displayed a high level of consistency is significant with regard to the validity of the GRS rating.
Figure 3.1 Distribution of mGRS scores for item “respect for tissue” in one-handed knot tying skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices ($p = 0.0138$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dots represent outliers. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 3.2 Distribution of mGRS scores for item “time and motion” in one-handed knot tying skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices ($p = 0.0001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outliers.
Figure 3.3 Distribution of mGRS scores for item “flow of operation” in one-handed knot tying skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices ($p = 0.0001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively.
In all of the individual items in the mGRS, the median scores would tend to confirm the expectation that experts score the highest in knot tying skill. The high score achieved by the experts would indicate that in general they have reached the competency level, as expected. The trainees’ scores were inferior to experts when it came to this skill. The median for the novices with regard to this domain indicates most novices scored significantly lower than trainees. The median would further confirm that there was an obvious difference in the scores in accordance with expectations from experience and literature (Paisley et al., 2001).

The ranges indicate that varieties in the scoring achieved by the experts, trainees and novices as given by the assessors. The range in the experts’ scores was in a narrow distribution because the experts would have a broadly similar level of competency based upon their experience. The trainee interquartile range and range would indicate that the scoring for these was somewhat broader and reflect the broad range of competency within the trainee group. The range and interquartile range of the novices was also tight and indicated that very few have any scoring that approached the level of the trainee.

Overall, the findings showed the participants’ skill in one-handed knot tying task with regards to “respect for tissue”, “time and motion” and “flow of operation” was consistent with their level of experience. This was to be expected as there is a great difference in the range of abilities of the experts, novices and the trainees. Therefore, the mGRS rating system delivered results that were in line with the expectations and this indicates that the mGRS is a valid assessment tool for one-handed knot tying skill.

3.4.2 Construct Validity of mGRS for Simple Interrupted Suturing Skill

Figures 3.4-3.9 showed the distribution of the scores for the six components of the mGRS for simple interrupted suturing skill. There was a significant difference in the all components of mGRS between experts, trainees and novices, (p<0.05). The experts consistently scored the highest mark, followed by the trainees and novices. This was expected based upon the literature (Scott
et al., 2007) and past experience. The experts score a good rating on average when it comes to suturing. The rating for the trainees show that many have suturing skills that are approaching the level of ability of experts. There is a wide range of ratings in this group which accounts for the differing levels of training and progress achieved by surgical trainees. The low scoring of the novices was also as expected. This would show that the evaluators are assessing and measuring what was expected, namely suturing skills of trainees.

![Box and whiskers plot showing distribution of mGRS scores for item “respect for tissue” in simple interrupted suturing skill.](image)

**Figure 3.4** Distribution of mGRS scores for item “respect for tissue” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p=0.0001). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 3.5 Distribution of mGRS scores for item “time and motion” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p=0.0001). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 3.6 Distribution of mGRS scores for item “instrument handling” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p=0.0001) The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively.
Figure 3.7 Distribution of mGRS scores for item “knowledge of instrument” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p<0.05). The boxes and whiskers represent interquartile range and range respectively.
Figure 3.8 Distribution of mGRS scores for item “flow of operation” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p=0.0001). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 3.9 Distribution of mGRS scores for item “knowledge of procedure” in simple interrupted suturing skill.

The box and whiskers plot shows a significant difference in this component of mGRS between experts, trainees and novices (p=0.0001). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
The rating for the suturing ability of the experts in all aspects of the skill was consistently higher than the other two groups. This was of a level that would be expected from experienced surgeons. The trainees showed generally a good level of skills, based upon the scores given by the assessors but inferior to the expert group. The novices’ competency in suturing skill was significantly lower and this finding would be expected of those who had little experience and only a basic knowledge of the area.

The individual items of mGRS tool have been proven to be able to measure scores of the participants and differentiate different level of experience in simple interrupted suturing skill. The scores provided by the assessors were as expected and it showed that there was a correlation between the competence level and experience levels.

The findings of these are broadly similar to the previous rating scores in one-handed knot tying skill. They broadly are what we would expect. This confirms the construct validity of the evaluative tool because they offered results that correlated to training and experience. Therefore, the mGRS tool is a valid assessment method for fundamental surgical skills.

### 3.4.3 Reliability of mGRS: One-handed Knot Tying Skill

The internal consistency of all the components in mGRS was $\alpha = 0.8436$. Previously published evidence suggests $\alpha > 0.70$ demonstrates adequate reliability (Acock, 2012). Table 3.2 demonstrates a weak item-test correlation of “respect for tissue” component with the total score of the three components in each trial, compared to “time and motion” and “flow of operation” components. By removing “respect for tissue” from mGRS, this inflates the $\alpha$ value to around 0.97. Apart from this, there was a high level of internal consistency shown in the rating and this is an indicator of reliability.
Table 3.2: The item-test correlation and α values for each item in mGRS for one-handed knot tying skill.

<table>
<thead>
<tr>
<th>Items</th>
<th>Item-test Correlation</th>
<th>Alpha (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>0.6802</td>
<td>0.9678</td>
</tr>
<tr>
<td>Time and motion</td>
<td>0.9457</td>
<td>0.6302</td>
</tr>
<tr>
<td>Flow of operation</td>
<td>0.9654</td>
<td>0.5781</td>
</tr>
</tbody>
</table>

This was a problem when it comes to the reliability of the GRS rating system. The internal consistency is essential for an assessment tool to be reliable and a figure such as that seen in the ‘respect for tissue’ item would suggest that there is some concern over reliability. The other two items scored very highly in contrast to ‘respect for tissue’.

3.4.4 Reliability of mGRS: Simple Interrupted Suturing Skill

The internal consistency of all the components in mGRS was α = 0.9804. Table 3.3 demonstrates a strong item-test correlation of all components of mGRS in each trial. By removing the tested component from mGRS, this inflates the α value to more than 0.90.
Table 3.3 The item-test correlation and α values for each item in mGRS for simple interrupted suturing skill.

<table>
<thead>
<tr>
<th>Items</th>
<th>Item-test Correlation</th>
<th>Alpha (α)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>0.8755</td>
<td>0.9869</td>
</tr>
<tr>
<td>Time and motion</td>
<td>0.9742</td>
<td>0.9739</td>
</tr>
<tr>
<td>Instrument handling</td>
<td>0.9741</td>
<td>0.9739</td>
</tr>
<tr>
<td>Knowledge of instrument</td>
<td>0.9764</td>
<td>0.9737</td>
</tr>
<tr>
<td>Flow of operation</td>
<td>0.9821</td>
<td>0.9739</td>
</tr>
<tr>
<td>Knowledge of procedure</td>
<td>0.964</td>
<td>0.9756</td>
</tr>
</tbody>
</table>

The ratings for each item were very similar. This strong correlation suggests that there was strong internal consistency with regard to the ratings and this is a good indicator of reliability.

3.4.5 Inter-rater Reliability: One-handed Knot Tying Skill

Table 3.6 and 3.7 showed inter-rater agreement, expressed as κ (kappa), between the 2 blinded assessors on each item. For intermediate κ values, Landis and Koch (1977) suggest the following interpretations (Table 3.4):
Table 3.4: The interpretations of Kappa, κ values according to Landis and Koch (1997).

<table>
<thead>
<tr>
<th>Kappa, κ values</th>
<th>Inter-rater agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0</td>
<td>Poor</td>
</tr>
<tr>
<td>0.00 – 0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81 – 1.00</td>
<td>Almost perfect</td>
</tr>
</tbody>
</table>

In order for the rating system to be reliable, there is a need for it to be at least a moderate degree of agreement between raters. A good level of agreement is needed to remove any doubts about the rating of an individual or group of raters being overly subjective. Implicit in this argument is that it is assumed that the experienced raters would generally be in agreement when scoring an item. A reasonable level of agreement among raters is important when considering the reliability of the GRS evaluation tool.

Table 3.5: Inter-rater agreement for each item in mGRS one-handed knot tying skill. K, kappa; p<0.005 statistically significant

<table>
<thead>
<tr>
<th>Components</th>
<th>Agreement</th>
<th>K</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>59.04%</td>
<td>0.4129</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>Time and Motion</td>
<td>64.46%</td>
<td>0.5249</td>
<td>p&lt;0.0001</td>
</tr>
<tr>
<td>Flow of operation</td>
<td>67.47%</td>
<td>0.5736</td>
<td>p&lt;0.0001</td>
</tr>
</tbody>
</table>

The assessors were in fair to moderate agreement on “respect for tissue” score but the agreement was stronger in “time and motion” item and “flow of operation” item. This indicates that the assessors generally evaluated the skills
of the experts, trainees and novices in the similar way. This is to prove that there is generally a good level of inter-rater reliability in the assessment of skills using the GRS rating system.

3.4.6 Inter-rater reliability: Simple interrupted suturing skill

The following section shows the inter-rater agreement in simple interrupted suturing skill. Table 3.6 showed inter-rater reliability t, expressed as κ (kappa), between the 2 blinded assessors on each item. The assessors were in fair agreement on “time and motion”, “instrument handling”, “knowledge of instrument” and “flow of operation” items. The level of agreement was only fair between the two raters. It was particularly low with ‘respect for tissue’ and “knowledge of procedure”. This highlights the limitation of the GRS system as there was a significant discrepancy between the two raters, especially on these two components in the rating scale.

<table>
<thead>
<tr>
<th>Components</th>
<th>Agreement</th>
<th>K</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td>33.78%</td>
<td>0.1076</td>
<td>0.0079</td>
</tr>
<tr>
<td>Time and motion</td>
<td>47.62%</td>
<td>0.2858</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Instrument handling</td>
<td>41.10%</td>
<td>0.2160</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Knowledge of instrument</td>
<td>39.19%</td>
<td>0.2053</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Flow of operation</td>
<td>47.30%</td>
<td>0.2987</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Knowledge of procedure</td>
<td>31.76%</td>
<td>0.1106</td>
<td>0.0036</td>
</tr>
</tbody>
</table>
3.5 Discussion

The GRS was modified to fit with the skills that were assessed. The excluded parameters were items that were not relevant to the skills and therefore, they were not being assessed in this study. For knot tying skill, three components were selected as the most relevant items to assess this skill, which were “respect for tissue”, “time and motion” and “flow of operation”. These parameters were selected to assess knot tying skill because the aim of the skill is to tie the surgical knot without damaging the tissues involved (such as pulling the tie off an artery by using excessive force) while carrying out the task in a timely efficient manner, as well as efficient in economy of movement. For suturing skill, it involves holding tissue together after an injury or an operation. The GRS was able to provide some reliable findings on the movements that are necessary for the skilful completion of a suturing procedure.

The relative tight ranges of the scores, especially the interquartile range would indicate that the rating was reliable. They show that the scores for experts are in a tight range showing that expert’s competency is generally similar. The trainees range was much broader indicating that a wide range of competency, typical of those who are still learning the skills and techniques of surgery.

In this chapter, we have demonstrated construct validity of the mGRS for the two skills we assessed. This means that it was measuring what it was supposed to be measuring. The mGRS was able to differentiate between experts, trainees and surgical novices. This is always an important criterion when it comes to evaluating any tool used for purposes of assessing skills.

In the two skills assessed, we demonstrated adequate reliability of the mGRS in terms of the internal consistency. Our results have shown Alpha, $\alpha$ value of 0.84 and 0.98 for the two skills respectively, which is higher than previously reported in the literature (Martin et al., 1997). This means that all the items in the mGRS measures the same construct in assessing the surgical skills.

As we have seen the mGRS rating for the different components was generally consistent. However, the “respect for tissue” component was something of an exception in the scale. The item as an individual parameter is not a reliable
item. It correlated poorly with the overall scale, especially in knot tying skill. The scale reliability was increased when this component was being removed from the scale. This is because “respect for tissue” did not fit well with the rest of the components in the scale. The rating of “respect for tissue” was an item that was accordingly difficult to evaluate objectively. This may cast some doubt upon the overall reliability of the GRS.

In addition, the inter-rater (IR) agreement is relatively low in judging the respect for tissue compared to other components. Inter-rater (IR) reliability is a psychometric property of a scale that refers to the extent to which ratings of the same performance by different observers are similar (Vassiliou et al., 2005). The differences in the GRS scores between the assessors remain a concern. The failure for the raters to agree on the score or to be as close as possible would indicate that personal perception and different experiences could influence the scoring. This is very much the case when the raters were only assessing without an awareness of the identity or profile of the experts. They based their ratings exclusively upon that which they witnessed in the video recording. Despite the ratings being carried out by experts with similar background, experience and standardised data collection, IR reliability was unable to reach the ideal agreement.

Subjectively, the two assessors could identify poor handling of tissue during knot tying by excessive pulling of the ligature that caused the rubber tubing to be markedly displaced from its original position. Even though the mGRS score uses Likert-type scale which scores 1 to 5 on each component, there is still variation among the assessors on how to score good or poor respect for tissue in knot tying skills. This raises the possibility that the view of one assessor could have distorted the findings. Therefore, in order to avoid personal bias, it is recommended that a minimum of two independent assessors are required.

A group of experts should utilise this observational tool when assessing a trainee either live during the assessment or via video recordings. When there are two or more observers, the scores ideally should not have a large discrepancy among them which means the inter-rater (IR) reliability should be high. An IR value of 0.8 means the assessors agreed in 80 percent of the scores but they were in disagreement in 20 percent of the scores. Inter-rater
agreement or reliability reported to be 0.67 which reflects significant differences of opinion of assessors in the subjective data they are evaluating (Olson et al., 2012). We would argue that having a high number of disagreements is not sufficient in assessing such a critical skill in surgery.

An assessment tool such as the GRS system is essential in determining if trainees have the competency to be able to perform surgical procedures independently. It is also important in the learning outcomes of trainees as the rating system is used to provide feedback on their progress. The feedback is used by both the trainees and trainers in order to identify any strengths and weaknesses in their performance. This is to improve their skills competency levels throughout their training years. Therefore, a good feedback based upon a valid and reliable objective scoring system is essential for successful training outcomes. The quality of the feedback will ensure that the trainees are clear in where they stand in terms of their skills proficiency and they are able to learn which areas they need to improve.

Therefore, it is necessary that there is the utmost confidence in the reliability of the tool. The GRS rating system based upon the research was shown to be reliable, but doubts remain. There are still issues with its ability to differentiate adequately between the skills levels of trainees, possible subjective judgements of the assessors and the assessors employing their own personal criteria when it came to assessing a component. This raises doubts about the complete reliability of the tool. Due to these problems and limitations with the current method of assessing surgical skills, a non-observational assessment tool could provide an adjunct role in assessing surgical skills. The use of both observational and non-observational assessment tools would provide a more reliable way of measuring skills rather than using one assessment tool in isolation. Therefore, there is an adjuvant role of a non-observational assessment tool in providing a reliable and consistent evaluation of surgical trainees' competency.
Chapter 4

Validity of a Motion Tracking System in Basic Surgical Skills
4.1 Introduction

Observer-depended scoring systems, for example OSATS, have a number of limitations. These include the fact that they are time consuming, labour intensive and open to a degree of subjectivity. These limitations have been previously discussed in Chapter 3. As a result there is a strong role for the additive effect of a more objective non-observer dependent assessment tool for the assessment of surgical skills.

Surgical specialties have initiated a trend towards a more objective and quantifiable measure of technical skill proficiency (Hayter et al., 2009). In minimally invasive surgery (laparoscopy and endoscopy), simulators have been developed with the ability to quantify the associated skills with specific metrics including total path length, movement efficiency and smoothness.

The use of motion analysis has been pioneered in gait analysis (Biryukova et al., 2000). It is used in tracking the movement of body parts. Its application is evident in various areas including sports such as golf, training an apprentice in spray painting and also in diagnostic simulators such as ultrasound simulation (http://polhemus.com/motion-tracking/case-studies#patriot). The surgical arena has used this technology to try and quantify surgical performance. Motion analysis allows assessment of surgical dexterity using parameters that are extracted from movement of the hands or laparoscopic instruments (van Hove et al., 2010). Surgical competencies, particularly in surgical trainees, can be determined by using these parameters.

Lord Ara Darzi and his researchers pioneered the use of an electromagnetic motion tracking device in surgery, called the Imperial College Surgical Assessment Device (ICSAD). This is a combination of a commercially available electromagnetic tracking system (Isotrack II, Polhemus Inc, Colchester, VT) and a bespoke computer software program (Datta et al., 2001). It measures the time taken, the number of movements, and the path length. All of these metrics have been shown to change with experience in laparoscopic surgery (Torkington et al., 2001) and in open surgery (bowel anastomosis and vein patch insertion) (Datta et al., 2001).
In this chapter, we further assessed the use of a novel electromagnetic tracking system in basic surgical skill tasks by using our own in-house computer software with a finger sensor. We used an off-the-shelf electromagnetic tracking system called the Patriot™ from Polhemus Inc, Colchester VT. The sensor was placed on the dorsum of right index finger. Our in-house software was designed to generate the classic metrics which are time and total path length (TPL). In addition new metrics were developed: average deviation distance from X-, Y- and Z-axis and average distance from centre of the bench model. The centre of the bench model is labelled as point of interest (POI) as we believe that hand motion is most efficient when the hands are at certain distance away from the centre of the workstation. Subjectively, when performing a certain task in open surgery such as tying surgical knots or suturing, a novice would have unnecessary movement of their hands which include moving their hands further away from the field of surgery. This is thought to be inefficient in view of the economy of the hand movement.

The aim of this chapter was to assess if these metrics were able to differentiate between groups with different surgical experience when performing basic surgical tasks. We also aimed to determine the relationship between these metrics and the observational assessment tool we used in the previous chapter, the OSATS mGRS.

4.2 Objectives

4.2.1 Hypothesis Underlying the Objectives

The metrics generated by our motion tracking system are relatively new compared to the published scores from the ICSAD. A rigorous analysis on these new metrics is required to determine the feasibility and validity of our new system. This is necessary as reliability and validity are essential in any evaluative tool.

The metrics should be able to differentiate the surgical skills according to the experience of the subjects. The motion tracking system should be able to provide data that reflects the experience and skills levels of different participants. In addition, the metrics should correlate with the items measured
by the observational assessment tool, OSATS mGRS which are currently the gold standard in assessing surgical skills.

4.2.2 Detailed Objectives

Objective 1: To demonstrate the construct validity of the metrics produced by our motion tracking system.

Objective 2: To determine the relationship between the metrics produced from the motion analysis system and the classical scoring system OSATS.

Objective 3: To determine the most efficient hand position by using the motion analysis system.

4.3 Materials and Methods

4.3.1 Recruitment of Participants

The medical students and all basic surgical trainees were contacted via general email. Posters were displayed in the Surgical Skills Laboratory in the National Surgical Training Centre (NSTC). Participation was voluntary. A full detail of the study was provided to the potential participants and they were encouraged to ask questions about the study.

Consultant surgeons who were involved in teaching surgical skills were invited to participate in this study. They were recruited as experts to determine the validity of the metrics and also to set the proficiency score according to each validated metric. They were regarded as experts based upon their years of training and experience in surgical procedures.

The participants were all informed about the nature of the study and what to expect during the course of the study. A written informed consent was then obtained from each participant to allow collection of data for research purposes only. All participants were advised that the information collected would be
stored and presented in an anonymous format. They were assured that all data and information would not be used for any other purpose than for the present study.

The assessments were performed in the National Surgical Training Centre, Royal College of Surgeons in Ireland.

### 4.3.2 Participant Demographics

Surgical trainees and medical students who expressed an interest in the study were asked to complete a simple questionnaire. The purpose of the questionnaire was to identify those participants who fulfilled the inclusion criteria and to exclude those who matched exclusion criteria as outlined in Chapter 2.

In total 34 surgical trainees and 27 medical students were initially recruited to the study. They were divided into two groups based on their experience: trainees and novices respectively. Each participant performed the surgical skill tasks, which were one-handed knot tying and simple interrupted suturing. All participants were required to perform these tasks for three times. The skills were assessed using the Patriot™ motion tracking system and the OSATS Global Rating Scale (GRS) via video recordings.

Consultant surgeons who fulfilled the inclusion criteria were included in the expert group. Each surgeon performed all the surgical skill tasks as described above and was assessed using the Patriot™ motion tracking system and the OSATS Global Rating Scale (GRS) via video recordings.

### 4.3.3 Skills Assessment Tool

The surgical skills for each participant were assessed using both an observational tool and a non-observational tool, as discussed in Chapter 2. A brief outline is provided below.

The observational tool used in this study was the OSATS global rating scale (GRS). Similar to the method in Chapter 3, the GRS items included in assessing
hand knot tying skill were “respect for tissue”, “time and motion” and “flow of operation”. The items included in assessing suturing skill were “respect for tissue”, “time and motion”, “instrument handling”, “knowledge of instrument”, “flow of operation” and “knowledge of procedure”. Each item was rated using Likert scale whereby 1 is poor performance and 5 is excellent performance. Two assessors who are consultant surgeons assessed the participants via anonymous video recordings during an independent session.

The Patriot™ motion tracking system was used as the non-observational tool. It consists of an electromagnetic source, a sensor attached to the right index finger and software. PiMgr software collected data according to the position of the sensor in terms of x-, y- and z-axis. Bespoke software was developed by RCSI to translate the raw data to simple metrics. These included total path length, average distance from the point of interest (POI) which was the centre of bench model and average distance from x-axis, y-axis and z-axis.

### 4.3.4 Surgical Skills Assessment

The skills that were assessed in this chapter were the one-handed knot tying skill and the simple interrupted suturing skill. Each subject was required to perform these skills three times in order to achieve an average of the three trials. Anonymous video recordings were taken during the assessment, as described in Chapter 2. Two experts (PN and DB) assessed the video recordings independently.

### 4.3.5 Statistical Analysis

Statistical analysis was performed using Stata 12 (StataCorp LP, Texas, USA). All data for each group were presented in box and whiskers plots which consist of the median, range and interquartile range. As mentioned in Chapter 2, the construct validity was assessed using Kruskal-Wallis test and Mann Whitney U test for further analysis. The Kruskal-Wallis test compares the median score across the groups (Acock, 2012). It is expressed as p value and it is included in the box and whiskers plots. The Mann-Whitney U test is used to compare
differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed (https://statistics.laerd.com/spss-tutorials/mann-whitney-u-test-using-spss-statistics.php).

In this chapter, the two independent groups are referred to the expert and novice groups. This test describes the results in standard scores or z-scores and most importantly, the significant level of the scores. If p<0.05, the score is considered as statistically significant.

4.4 Results

4.4.1 Participant Demographics

A total of 28 trainees and 25 novices were included in hand knot tying skill and 35 trainees and 27 novices were included in suturing skill (Table 4.1). Five experts were excluded from hand knot tying skill because they formed the loops with their right hand. Six trainees and two novices were excluded due to technical error with the video recordings.

The age of all participants in the expert group was greater than 40 and in the novice groups ranged between 19 and 25. The age in the trainee group ranged between 26 and 40. Figure 4.1 shows the number of trainees in different years of training in the National Surgical Training Programme.
Table 4.1 Participant demographics for hand knot tying and suturing skills.

<table>
<thead>
<tr>
<th></th>
<th>Expert</th>
<th>Trainee</th>
<th>Novice</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hand Knot Tying Skill</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td><strong>Dominant Hand</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>5</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>Left</td>
<td>0</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><strong>Suturing Skill</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>Female</td>
<td>1</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td><strong>Dominant Hand</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right</td>
<td>5</td>
<td>34</td>
<td>24</td>
</tr>
<tr>
<td>Left</td>
<td>5</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>
Figure 4.1 Number of trainees who participated according to their year of training.
4.4.2 Construct Validity of the Metrics in One-handed Knot Tying Skill

The median values of the various metrics for the three different groups are displayed in Table 4.2. There was a significant difference found between the groups for time, total path length, distance from the point of interest (POI) and deviation distance from the Z-axis. The significant results are shown in box & whisker plot format in Figures 4.2-4.5.

Table 4.2 Metrics of the three different experience groups. Values displayed are medians ± standard deviation, with the range in parentheses. P values were calculated using k-wallis test. ‡ POI, Point of interest indicates the centre of bench model. * p<0.05 is statistically significant.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Experts (n=5)</th>
<th>Trainees (n=28)</th>
<th>Novices (n=25)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>19.9 ± 5.4</td>
<td>22.9 ± 8.2</td>
<td>81.2 ± 30.9</td>
<td>0.0001*</td>
</tr>
<tr>
<td></td>
<td>(9.7-22.6)</td>
<td>(9.7-44.8)</td>
<td>(47.3-163.7)</td>
<td></td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>126.5 ± 31.7</td>
<td>132.9 ± 41.6</td>
<td>336.8 ± 142.3</td>
<td>0.0001*</td>
</tr>
<tr>
<td></td>
<td>(96.7-176.0)</td>
<td>(71.5-255.0)</td>
<td>(154.7-732.5)</td>
<td></td>
</tr>
<tr>
<td>Distance from POI‡</td>
<td>4.1 ± 1.0</td>
<td>5.1 ± 1.2</td>
<td>8.1 ± 1.3</td>
<td>0.0001*</td>
</tr>
<tr>
<td></td>
<td>(3.9-6.3)</td>
<td>(3.1-8.1)</td>
<td>(3.6-10.4)</td>
<td></td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>2.5 ± 0.4</td>
<td>2.3 ± 0.7</td>
<td>2.7 ± 1.0</td>
<td>0.2066</td>
</tr>
<tr>
<td></td>
<td>(1.9-3.0)</td>
<td>(1.4-4.9)</td>
<td>(1.2-5.4)</td>
<td></td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>0.9 ± 0.5</td>
<td>0.9 ± 0.4</td>
<td>1.2 ± 0.2</td>
<td>0.0708</td>
</tr>
<tr>
<td></td>
<td>(0.7-1.8)</td>
<td>(0.5-2.4)</td>
<td>(0.7-1.5)</td>
<td></td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>1.2 ± 0.3</td>
<td>1.3 ± 0.4</td>
<td>2.0 ± 0.5</td>
<td>0.0003*</td>
</tr>
<tr>
<td></td>
<td>(1.0-1.6)</td>
<td>(0.8-2.8)</td>
<td>(0.9-3.2)</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.2 Distribution of average time taken to complete one-handed knot tying task.

The box and whiskers plot shows a significant difference between experts, trainees and novices (p < 0.001). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 4.3 Distribution of the total path length (TPL) between the three subject groups.

The box and whiskers plot shows a significant difference between experts, trainees and novices (p < 0.001). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents an outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 4.4 Distribution of the average distance from the point of interest (POI) or centre of the bench model between the three subject groups.

The box and whiskers plot shows a significant difference between experts, trainees and novices ($p < 0.001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 4.5 Distribution of the average distance from the Z-axis between the three subject groups.

The box and whiskers plot shows a significant difference between experts, trainees and novices (p < 0.001). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dots represent outliers. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Table 4.3 The Z scores and p values from Mann Whitney U test for each metrics when compared between the experts and novices. ‡ POI, Point of interest indicates the centre of bench model. * p<0.05 is statistically significant.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Z score</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>-3.504</td>
<td>0.0005*</td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>-3.451</td>
<td>0.0006*</td>
</tr>
<tr>
<td>Distance from POI‡ (mm)</td>
<td>-3.244</td>
<td>0.0012*</td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>-0.597</td>
<td>0.5506</td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>-0.493</td>
<td>0.6220</td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>-2.673</td>
<td>0.0075*</td>
</tr>
</tbody>
</table>

The Mann Whitney U test was conducted to determine the difference in two independent groups which were the experts and the novices. The metrics of time, total path length, distance from POI and deviation distance from Z-axis showed statistically significant difference in scores between the experts and novices. This means that these novel metrics from the Patriot are valid to be used in one-handed knot tying skill.
4.4.3 Construct Validity of the Metrics in Simple Interrupted Suturing Skill

The median values of the various metrics for the three different groups are displayed in Table 4.4. There was a significant difference found between the groups for time, total path length, distance from the point of interest (POI) and deviation distance from the Z-axis. The significant results are shown in box & whisker plot format in Figures 4.6-4.9.

Table 4.4 Metrics of the three different experience groups. Values displayed are medians ± standard deviation, with the range in parentheses. P values were calculated using k-wallis test.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Experts (n=10)</th>
<th>Trainees (n=35)</th>
<th>Novices (n=27)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>24.6 ± 4.6 (13.2-29.4)</td>
<td>33.9 ± 15.6 (21.0-90.0)</td>
<td>75.2 ± 32.6 (40.7-177.3)</td>
<td>0.0001*</td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>99.5 ± 30.4 (73.9-164.3)</td>
<td>164.7 ± 34.9 (90.6-224.5)</td>
<td>190.1 ± 82.4 (129.0-472.1)</td>
<td>0.0001*</td>
</tr>
<tr>
<td>Distance from POI‡</td>
<td>2.7 ± 0.4 (2.1-3.8)</td>
<td>2.6 ± 0.3 (1.9-3.4)</td>
<td>2.8 ± 0.3 (2.2-3.5)</td>
<td>0.0145*</td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>1.4 ± 0.5 (0.6-1.9)</td>
<td>1.3 ± 0.4 (0.7-2.3)</td>
<td>1.6 ± 0.4 (0.7-2.7)</td>
<td>0.3311</td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>1.6 ± 0.5 (0.9-2.8)</td>
<td>1.7 ± 0.3 (1.2-2.5)</td>
<td>1.8 ± 0.3 (1.4-2.5)</td>
<td>0.2082</td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>0.8 ± 0.3 (0.7-1.7)</td>
<td>1.1 ± 0.2 (0.5-1.5)</td>
<td>1.2 ± 0.3 (0.7-2.2)</td>
<td>0.0193*</td>
</tr>
</tbody>
</table>

‡ POI, Point of interest indicates the centre of bench model
* p<0.05, statistically significant
The box and whiskers plot shows a significant difference between experts, trainees and novices ($p < 0.001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 4.7 Distribution of the total path length (TPL) between the three subject groups.

The box and whiskers plot shows a significant difference between experts, trainees and novices ($p < 0.001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
The box and whiskers plot shows a significant difference between experts, trainees and novices ($p < 0.001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Figure 4.9 Distribution of the average distance from the Z-axis between the three subject groups.

The box and whiskers plot shows a significant difference between experts, trainees and novices ($p < 0.001$). The horizontal lines within boxes are the median. The boxes and whiskers represent interquartile range and range respectively. The dots represent outliers. The interquartile range may be a more reliable measurement of the range as it removes any outliers who may distort the scores.
Table 4.5 The Z scores and p values from Mann Whitney U test for each metrics when compared between the experts and novices. ‡ POI, Point of interest indicates the centre of bench model. * p<0.05 is statistically significant.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Z-score</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (s)</td>
<td>-4.617</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>-4.138</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Distance from POI‡ (mm)</td>
<td>-1.129</td>
<td>0.2591</td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>-1.060</td>
<td>0.2891</td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>-1.231</td>
<td>0.2183</td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>-2.565</td>
<td>0.0103*</td>
</tr>
</tbody>
</table>

The Mann Whitney U test was conducted to determine the difference in two independent groups which were the experts and the novices. The metrics of time, total path length and deviation distance from Z-axis showed statistically significant difference in scores between the experts and novices. This means that these novel metrics from the Patriot™ motion tracking system are valid to be used in simple interrupted suturing skill.
4.4.4 Concurrent Validity of the Metrics in One-handed Knot Tying Skill

The relationship between the metrics from the Patriot™ motion tracking system and the participants’ OSATS mGRS scores in one-handed knot tying skill are demonstrated in Table 4.6 (Spearman’s rank correlation). The results were all found to be statistically significant except for correlation between the deviation distance from x-axis, “respect for tissue” and total OSATS mGRS score.

The total OSATS mGRS scores were highly correlated with time, total path length (TPL), deviation distance from Z-axis and from point of interest (POI) or centre of the rubber tubing on knot tying jig. There was a high correlation found between total path length and all three OSATS mGRS item individually. The deviation distance from Z-axis and from POI showed good correlations with the OSATS mGRS items “time and motion” and “flow of operation”.
Table 4.6 Correlation between the OSATS mGRS scores and the metrics from Patriot™ motion tracking system in one-handed knot tying skill. The values indicate the Spearman’s correlation coefficients or rho ($\rho$) which measures the strength of association between the mGRS scores and the metrics from Patriot™. The association is stronger if the correlation coefficient is closer to -1 or +1.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Correlation with each OSATS mGRS items</th>
<th>Correlation with total GRS scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Respect for tissue</td>
<td>Time and motion</td>
</tr>
<tr>
<td>Time</td>
<td>-0.5319</td>
<td>-0.9488</td>
</tr>
<tr>
<td>Total Path Length</td>
<td>-0.6060</td>
<td>-0.8424</td>
</tr>
<tr>
<td>Distance from POI‡</td>
<td>-0.4010</td>
<td>-0.6890</td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>-0.1662*</td>
<td>-0.3008</td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>-0.3940</td>
<td>-0.2757</td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>-0.3719</td>
<td>-0.4813</td>
</tr>
</tbody>
</table>

* $p \geq 0.05$, statistically not significant, ‡ POI, Point of interest indicates the centre of bench model
4.4.5 Concurrent Validity of the Metrics in Simple Interrupted Suturing Skill

Table 4.7 shows the Spearman’s rank correlation between the metrics from the Patriot™ motion tracking system and participants’ GRS scores in simple interrupted suturing skill, expressed as rho values. The total GRS scores and each GRS item showed good correlation with time and total path length (TPL). There was significant correlation between deviation distances from Z-axis, point of interest (POI), total GRS scores and all GRS items except “respect for tissue” item.
Table 4.7 Correlation between the OSATS mGRS scores and the metrics from Patriot™ motion tracking system in simple interrupted suturing skill. The values indicate the Spearman’s correlation coefficients or rho (ρ) which measures the strength of association between the mGRS scores and the metrics from Patriot™. The association is stronger if the correlation coefficient is closer to -1 or +1.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Correlation with each GRS items</th>
<th>Correlation with total GRS scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Respect for tissue</td>
<td>Time motion and Instrument handling</td>
</tr>
<tr>
<td>Time</td>
<td>-0.4137</td>
<td>-0.8728</td>
</tr>
<tr>
<td>Total Path Length</td>
<td>-0.4329</td>
<td>-0.6494</td>
</tr>
<tr>
<td>Distance from POI‡</td>
<td>0.0871*</td>
<td>-0.0882*</td>
</tr>
<tr>
<td>Deviation distance from X-axis</td>
<td>0.0586*</td>
<td>0.0059*</td>
</tr>
<tr>
<td>Deviation distance from Y-axis</td>
<td>0.1284*</td>
<td>-0.1274*</td>
</tr>
<tr>
<td>Deviation distance from Z-axis</td>
<td>-0.2274*</td>
<td>-0.3037</td>
</tr>
</tbody>
</table>

* p≥0.05, statistically not significant, ‡ POI, Point of interest indicates the centre of bench model
4.5 Discussion

Motion analysis is widely used in sports and training spray painters (http://polhemus.com/motion-tracking/all-trackers/patriot/). Electromagnetic tracking has the advantage of not requiring line of sight, which is necessary for any visual or infrared-based system (Bann et al., 2003). This permits recording of dexterity in a 3-D forms.

Lord Darzi and his research team devised the Imperial College Surgical Assessment Device (ICSAD) in 1999 that uses an electromagnetic motion tracking system. This motion analysis device uses an alternating current electromagnetic system with passive receiver attached to the dorsum of the hand over the mid-shaft of the third metacarpal (Bann et al., 2003). ICSAD is able to quantify surgical skills by measuring the total path length, time and number of hand movements. Mason et al. (Mason et al., 2013) have reviewed the published evidence as it relates to motion analysis and the assessment of surgical performance. This systematic review reported construct validity of ICSAD and other forms of motion analysis devices such as ProMIS Augmented Reality Simulator and Hiroshima University Endoscopic Surgical Assessment Device (HUESAD) in assessing laparoscopic skills.

In this study, we used an electromagnetic tracking device called Patriot™ from Polhemus Inc. Instead of attaching the sensor on to the dorsum of hand in ICSAD, the sensor was fixed to the dorsum of right index finger of the subjects with Velcro® strap at the level of proximal phalanx. This is thought to give a more precise reading. RCSI developed bespoke software to translate the raw data from the PiMgr software into total path length (TPL), average distance from point of interest (POI), average deviation distance from the X-, Y- and Z-axis. These parameters were chosen to assess whether surgeons utilised an optimum 3-D distance when performing tasks such as suturing. This is a novel concept in surgical assessment of technical skills.

Our results have demonstrated construct validity for both the one-handed knot tying task and the simple interrupted suturing skill for the metrics time, total path length, point of interest and deviation from the Z-axis. The novel parameters
were able to differentiate subjects according to level of experience along with the validated metrics as reported in literature (Ezra et al., 2009, Datta et al., 2001). This implies that a surgical novice moved his or her hand further away from the virtual Z-axis and midpoint of the work station than experts or surgical trainees, as seen subjectively in the video recordings. Therefore, it is postulated that this pattern of movement is less efficient. The lack of significant change in X- and Y-axis may reflect the standard suture tie length used in this experiment. This was limiting the movement in these axes.

The second objective of our study was to investigate the relationship between the Patriot motion tracking system as a non-observational tool and the observational tool (GRS). For the one-handed knot tying skill our results demonstrated a significant correlation between all the metrics generated by the Patriot and the items of the GRS scoring tool. The only parameter that failed to demonstrate a significant relationship was deviation from the x-axis and “respect for tissue”. For the simple interrupted suturing skill we found a significant correlation between time, total path length and deviation from the z-axis and the total GRS score. The results for the suturing skill were less consistent when compared to the knot tying skill. This may be explained by the fact that a greater number of items were included in the GRS for the suturing skill. The “respect for tissue” item as we have previously demonstrated in Chapter 3 can be a very subjective parameter. This is reflected in the poor inter-rater reliability. In both skills, the metrics correlated well with the GRS items especially items involving motion and flow of operation. We could safely suggest that the Patriot provides more objective score than the observer-dependant scale.

However, the metrics from Patriot motion tracking system failed to show a more convincing correlation with the scale assessing tissue handling, even though the values were statistically significant. It also did not incorporate end-product errors such as slip knots and gapping of the skin edges. Hand-tracking data appear to confirm that skilled individuals demonstrate a shorter path length, make fewer movements, and take less time to perform an operation, but with the caveat that this improved performance is not accompanied by an increase in errors (Gallagher et al., 2013).
The motion tracking device is not able to detect surgical errors, unlike VR simulators such as LapSim and LapMentor. These simulators are programmed to identify surgical errors and incorporate it with other metrics. We suggest that this may be overcome by adding scores from assessment of the knot quality with force gauge device. It also highlights the fact that we as yet do not have a single solution for skill assessment. At present, due to the complexity of this problem, a multimodality approach remains in place. As the technology continues to develop, surgical educators should move away from using the traditional observational methods but instead incorporating more technology that gives objective measurements when assessing trainees in this high-stakes profession.
Chapter 5
The application of Patriot™ motion tracking system in a proficiency-based assessment of basic surgical skills
5.1 Introduction

Work-hour limitations, financial constraints, and ethical concerns have forced educators to explore new options for increasing the efficiency of training and teaching outside of the operating room (Darosa et al., 2003). Laparoscopy and endoscopy training are leading in ensuring a validated curriculum is implemented to hone the skills among surgical trainees. There is a steep learning curve in these skills requiring a focused practice and assessment in the simulation laboratory.

Proficiency-based training has been described as learning environments in which the trainee progresses from less to more technically demanding skills and tasks only after achieving predefined criteria (Stefanidis et al., 2006) (Aggarwal et al., 2006). One widely available simulation-based assessment and certification program is the Fundamentals of Laparoscopic Surgery (FLS) developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) and now administered by SAGES and the American College of Surgeons (Vassiliou and Feldman, 2011). The FLS program incorporates tasks from the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) program, including laparoscopic suturing, and uses well-described, low-fidelity inanimate models (Scott et al., 2007). The proficiency scores were determined by a group of experts in the skills and the trainees or users are required to reach these predetermined scores before they could proceed to the next level or task. These proficiency scores act as an aim for the trainees to achieve and subsequently motivate them to keep practising until a high standard of surgical skills is accomplished. In order to do this, an automated objective measurement is much desirable, such as simulators in laparoscopy or endoscopy, as it does not require any expert surgeons to monitor and assess the performance.

However, there is a paucity of literature that describes the use of objective measurement in proficiency-based curriculum for open surgical skills. Scott et al developed a cost-effective open skills curriculum that included expert-derived proficiency goals based on time and error, similar to FLS programme (Scott et al., 2007). This requires observation from proctors to assess the errors and time
taken to complete the tasks including knot tying and suturing skill. This is obviously labour intensive and not readily available for the trainees to continue practising in their own time, unlike in laparoscopy or endoscopy training.

Previously in Chapter 4, we have validated automated metrics from Patriot™ hand tracking system rigorously. These validated metrics can be regarded more superior than some of the elements in the Global Rating Scale (GRS) from the Objective Structured Assessment of Technical Skill (OSATS). This was inferred after successfully proving the concurrent validity of the automated metrics that could measure the same components from the GRS without depending on experts or proctors’ observation. We have suggested that these metrics should be fully utilised in assessing open surgical skills rather than an observer-dependant assessment tool.

In this chapter, we applied the concept of proficiency-based training by using the validated metrics from the Patriot™ motion tracking system. We determined the proficiency goals for each of these metrics in knot tying and suturing skills. The performance of surgical trainees in Year 1 and Year 2 of the surgical training programme were assessed using this method. Their scores were then analysed against these predetermined proficiency goals. Our intention was to have an objective automated tool that can be integrated into the national training curricula as part of the training module. This will help the trainees to practise and eventually achieve the desired performance in the most fundamental skills in surgery.

5.2 Objectives

5.2.1 Hypothesis Underlying the Objectives

The Patriot™ motion tracking system is able to provide a set of automated metrics in order to evaluate the performances in surgical technique. It does not require a group of assessors to observe trainees during the assessment. The trainees would be able to use this system in their own time to practice towards proficiency.
The Patriot™ also can be used as an assessment tool in proficiency-based assessment. It can be used to monitor trainees’ progress in the training centre and identify which trainees that require extra attention on their technical skills.

5.2.2 Detailed Objectives

Objective 1: To evaluate the use of motion tracking system in proficiency-based assessment of basic surgical skills

Objective 2: To map out the progress of surgical trainees using the metrics in motion tracking system

Objective 3: To use the motion tracking metrics in recognizing trainees’ competency in their surgical skills.

5.3 Materials and Methods

5.3.1 Recruitment of Participants

As described in Chapter 2 and 3, all basic surgical trainees were contacted via general email and posters were displayed in the Surgical Skills Laboratory in the National Surgical Training Centre (NSTC). Participation was voluntary and informed consent obtained. Data were processed in an anonymous format. All the assessment was performed in the National Surgical Training Centre, Royal College of Surgeons in Ireland.

The participants were grouped into two cohorts for this chapter. The first cohort (Cohort A) consisted of surgical trainees who were in the middle of their basic surgical training programme (BST Year 1 and Year 2). The second cohort (Cohort B) was the surgical trainees who were recruited into the new training programme. In the new training programme, these trainees were required to attend a 5-day boot camp before they started their clinical placement in different hospitals.

5.3.2 Surgical Skills Assessment

The skills that were assessed in this chapter were one-handed knot tying skill and simple interrupted suturing skill. The detailed outline of these tasks was
discussed in Chapter 2. The trainees’ performance was measured using the Patriot™ motion tracking system, which is outlined below.

The targeted proficiency level was determined by using the score of the consultant surgeons in Chapter 3. We calculated the mean and standard deviation of the scores. The formula used to calculate the targeted proficiency level is the mean score of the experts plus 1 standard deviation (Table 5.1).

Both cohorts were asked to perform the two simple fundamental skill tasks three times and the averages of their performance were calculated to reduce bias. The average scores for Cohort A were plotted around the target proficiency level calculated in Chapter 3. This will reveal which trainees have or have not achieved the desired proficiency level of the skillset.

Trainees’ skills in Cohort B were assessed prior to the mandatory boot camp and after five months in the training programme. The scores were compared in order to determine the retention of skill throughout the training using the automated assessment tool.

### 5.3.3 Skills Assessment Tool

Patriot™ motion tracking system was used to assess the skills performance, as described in Chapter 2 and 3. The metrics used in this chapter for one-handed knot tying skill include time, total path length (TPL), distance from z-axis and POI. As for simple interrupted suturing skill, the metrics were time, TPL, distance from z-axis and average distance from the point of interest (POI) which was the centre of bench model. These metrics were validated in Chapter 3.

### 5.3.4 Statistical Analysis

Statistical analysis was performed using Stata 12 (StataCorp). Frequency counts were used to determine trainees who reached proficiency and did not reach proficiency target. The metrics and GRS scores have normal distributions according to Shapiro-Wilk W test. The paired t-test was used to analyse the metrics including the GRS before and five months after training. We used the Pearson product-moment correlation coefficient to compare the correlation between the metrics and the OSATS score.
Table 5.1 The range, mean and standard deviation of the experts’ scores for one-handed knot tying skill and suturing skill. *(POI = Point of interest or centre of the bench model)*

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Range</th>
<th>Mean</th>
<th>Standard Deviation (SD)</th>
<th>Proficiency Level (Mean + SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-handed knot tying skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (s)</td>
<td>9.7-22.6</td>
<td>18.2</td>
<td>5.4</td>
<td>23.6</td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>96.7-176.0</td>
<td>129.0</td>
<td>31.7</td>
<td>160.7</td>
</tr>
<tr>
<td>Deviation from Z-axis (mm)</td>
<td>1.0-1.6</td>
<td>1.3</td>
<td>0.3</td>
<td>1.6</td>
</tr>
<tr>
<td>Distance from POI* (mm)</td>
<td>3.9-6.3</td>
<td>4.6</td>
<td>1.0</td>
<td>5.6</td>
</tr>
<tr>
<td><strong>Simple interrupted suturing skill</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time (s)</td>
<td>13.2-29.5</td>
<td>23.5</td>
<td>4.6</td>
<td>28.1</td>
</tr>
<tr>
<td>Total Path Length (mm)</td>
<td>77.7-164.3</td>
<td>113.2</td>
<td>30.4</td>
<td>143.6</td>
</tr>
<tr>
<td>Distance from Z-axis (mm)</td>
<td>0.7-1.7</td>
<td>0.9</td>
<td>0.3</td>
<td>1.2</td>
</tr>
<tr>
<td>Distance from POI* (mm)</td>
<td>2.1-3.8</td>
<td>2.7</td>
<td>0.4</td>
<td>3.1</td>
</tr>
</tbody>
</table>
5.4 Results

5.4.1 Proficiency-based assessment of skills

One-handed Knot Tying Skill

Figure 5.1-5.4 showed the performances of surgical trainees in one-handed knot tying skill (n=34). A proficiency level was drawn as dashed line for each metric.

![Figure 5.1](image)

**Figure 5.1** The time taken to complete one-handed knot tying skill task of all the trainees in the study (n=34).

The dashed line represents the proficiency level of 23.6s as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who took shorter time than the time at proficiency (n=21, 62%). The round shape points above the dashed line are the trainees who took longer time to complete the task and did not reach proficiency level (n=13, 38%).
Figure 5.2 The total path length (TPL) of all the trainees in the study who performed one-handed knot tying skill.

The dashed line represents the proficiency level of 160.7mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who have shorter path length and considered as proficient in their skill (n=25, 74%). The round shape points above the dashed line are the trainees with longer path length and did not reach the proficiency level (n=9, 26%).
Figure 5.3 The distance from the Z-axis of all the trainees in the study who performed one-handed knot tying skill.

The dashed line represents the proficiency level of 1.6 mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who were considered as proficient in their skill (n=26, 77%). The round shape points above the dashed line are the trainees who did not reach proficiency level (n=8, 23%).
Figure 5.4 The distance from the POI of all the trainees in the study who performed one-handed knot tying skill.

The dashed line represents the proficiency level of 5.6mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who were closer to the POI and considered as proficient in their skill (n=25, 74%). The round shape points above the dashed line are the trainees who were further from the POI and did not reach proficiency level (n=9, 26%).
Simple Interrupted Suturing Skill

**Figure 5.5-5.8** showed their performance in simple interrupted suturing skill (n=35). A proficiency level was drawn as dashed line for each metric.

Figure 5.5 The time taken by all the trainees in the study to complete the simple interrupted suturing skill task.

The dashed line represents the proficiency level of 28 seconds as shown in **Table 5.1**. The diamond shape points below the dashed line are the trainees who took shorter time than the proficiency level (n=9, 26%). The round shape points above the dashed line are the trainees with took longer time to complete the task and did not reach proficiency level (n=26, 74%).
Figure 5.6 The total path length (TPL) of all the trainees in the study who performed simple interrupted suturing skill.

The dashed line represents the proficiency level of 143.6mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who have shorter path length and considered as proficient in their skill (n=13, 52%). The round shape points above the dashed line are the trainees with longer path length and did not reach proficiency level (n=22, 48%).
Figure 5.7 The distance from POI of all the trainees in the study who performed simple interrupted suturing skill.

The dashed line represents the proficiency level of 1.2mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who were considered as proficient in their skill (n=26, 74%). The round shape points above the dashed line are the trainees who did not reach proficiency level (n=9, 26%).
Figure 5.8 The distance from point of interest (POI) of all the trainees in the study who performed simple interrupted suturing skill.

The dashed line represents the proficiency level of 3.1mm as shown in Table 5.1. The diamond shape points below the dashed line are the trainees who were closer to the POI and considered as proficient in their skill (n=32, 91%). The round shape points above the dashed line are the trainees who were further from the POI and did not reach proficiency level (n=3, 9%).

*POI = Point of interest or centre of the bench model
5.4.2 Monitoring trainees’ progress

Total of 20 trainees were recruited in Cohort B and 14 of them (return rate 70%) returned for one-handed knot tying task and 16 (return rate 80%) for simple interrupted suturing task after five months of training for the second part of this chapter.

**One-handed Knot Tying Skill**

*Figure 5.9-5.12* showed the difference in time, total path length (TPL), the distance from the Z-axis and point of interest (POI) before and after five months of training.

The time taken to complete the task at the beginning of the training programme was (20.0 +/- 7.1 seconds) and after five months of training, the time taken was significantly shorter (31.5 +/- 17.3 seconds) with p value of 0.0068. This showed that the trainees were becoming faster in tying surgical knots after almost half way through their first year of training.

The TPL also showed statistically significant difference in the trainees’ performance between baseline and after five months of training. The TPL at baseline was 168.88 +/- 66.43mm and this was decrease to 139.15 +/- 38.68mm at five-month assessment. The TPL was significantly shorter by 29.73mm (p=0.0085, ie p < 0.05). The trainees in this cohort have shown that they used shorter path length to complete this task which implies more economical hand motion.

The mean distance from Z-axis metric at baseline was 1.65 +/- 0.71mm and after five months in training, the mean distance was 1.46 +/- 0.55mm. This metric was shorter by 0.19mm, although this is not statistically significant (p=0.2289).

The mean distance from POI metric at baseline was 4.72 +/- 1.06mm and after five months in training, the mean distance was 5.46 +/- 0.97mm. This metric was longer by 0.74mm (p=0.024, p < 0.05).
Figure 5.9 Difference in the time taken to complete the one-handed knot tying task at the start of training (Month 0 or baseline) and at five months of training. (Shaharan et al., 2015)

The box and whiskers plot showed a statistically significant difference in the time taken before and after five months of training (p<0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier.
Figure 5.10 Difference in the total path length (TPL) at the start of training (Month 0 or baseline) and at five months of training. (Shaharan et al., 2015)

The box and whiskers plot showed a statistically significant difference in the TPL before and after five months of training (p<0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 5.11 Difference in the distance from Z-axis at the start of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot did not show statistically significant difference in the distance from Z-axis before and after five months of training (p>0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 5.12 Difference in the deviation distance from the point of interest (POI) at the beginning of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot showed a statistically significant difference in the deviation distance from the POI before and after five months of training (p<0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier.

*POI = Point of interest or centre of the bench model
Simple Interrupted Suturing Skill

**Figure 5.13-5.16** showed the difference in time, total path length (TPL), the distance from the Z-axis and point of interest (POI) before and after five months of training.

The time taken to complete the task at the beginning of the training programme was (45.9 +/- 11.2 seconds) and after five months of training, the time taken was shorter (38.7 +/- 17.1 seconds) with p value of 0.1193. Although the difference was not statistically significant, this showed that the trainees were becoming faster in tying surgical knots after almost half way through their first year of training.

The TPL at baseline was 165.42 +/- 43.13mm and after five months of training, the TPL was 160.74 +/- 34.77mm at five-month assessment. The TPL was shorter by 4.68mm only (p=0.6355, ie p > 0.05). The trainees in this cohort did not show any significant improvement in this task.

The mean distance from Z-axis metric at baseline was 1.19 +/- 0.21mm and after five months in training, the mean distance was 1.10 +/- 0.27mm. This metric was shorter by 0.09mm, although this is not statistically significant (p=0.1479).

The mean distance from POI metric at baseline was 2.62 +/- 0.38mm and after five months in training, the mean distance was 2.68 +/- 0.28mm. This metric was shorter only by 0.06mm and this is not statistically significant. (p=0.6175, p > 0.05).
Figure 5.13 Difference in the time taken to complete the task at the beginning of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot did not show statistically significant difference in the time before and after five months of training ($p>0.05$). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively. The dot represents outlier.
Figure 5.14 Difference in the total path length (TPL) to complete the task at the beginning of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot did not show statistically significant difference in the TPL before and after five months of training (p>0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
Figure 5.15 Difference in the deviation distance from the Z-axis to complete the task at the beginning of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot did not show statistically significant difference in the deviation distance from the Z-axis before and after five months of training (p>0.05). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively.
**Figure 5.16** Difference in the deviation distance from the point of interest (POI) at the beginning of training (Month 0 or baseline) and at five months of training.

The box and whiskers plot did not show statistically significant difference in the deviation distance from the POI before and after five months of training ($p>0.05$). The horizontal lines within boxes are median. The boxes and whiskers represent interquartile range and range respectively. The dots represent outliers.
5.5 Discussion

The use of motion tracking and analysis in assessing surgical skills has been described mainly in laparoscopic skills (Neary et al., 2008, Smith et al., 2002). There is a lack of literature that describes the application of such technology in assessing open surgical skills. A group of researchers in Imperial College London developed the Imperial College Surgical Assessment Device (ICSAD) that utilises the concept of motion tracking in open surgical skills (Datta et al., 2001). Previous studies established the construct validity of ICSAD on open surgical procedures including small bowel anastomosis, vein patch insertion (Datta et al., 2002) and ophthalmic micro-suturing (Saleh et al., 2006). Similar to the ICSAD, we used the Patriot™ device which is an off-the-shelf apparatus from Polhemus Inc. (Colchester, VT, USA) that tracks full six degree of freedom motion. The device is widely used in motion analysis on areas of simulation and training such as training painters and ultrasound simulators (http://polhemus.com/motion-tracking/all-trackers/patriot/).

The skills that we assessed were fundamental skill in surgery. The surgical trainees learn these important skills at an early stage and they are expected to be proficient before they can proceed to perform simple procedures such as excision of skin or subcutaneous lesion or more complex procedures such as repair of tendon or nerve. The trainees would require direct guidance and abundance of practice in order to be proficient in these skills, as the saying goes “practice makes perfect”. By having an expert or supervisor to observe them during practice session is not feasible when clinical work takes priority. Therefore, an automated system such as the Patriot™ system would be necessary to allow the trainees to practice and record their performance in their own time.

In this chapter, we have shown the potential application of this device in surgical training. The Patriot™ system tracked the hand motion of the user when performing the tasks and the raw data was converted to meaningful objective scores. After completing the tasks, the users were provided with their scores that mirrored their performance. With this, we were able to assess the level of proficiency on these trainees at any given time. Their scores were compared
with the set goals or predetermined proficiency levels. These levels were
determined by a group of senior surgeons in our college and the scores were
calculated as follow:-

| Proficiency level = mean score of the expert surgeons + 1 standard deviation |

According to our results, majority of the first and second year trainees in this
cohort were proficient at one-handed knot tying skills as expected and less than
one third of them did not reached proficiency. In simple interrupted suturing skill,
most of the trainees have reached the proficiency level in terms of the path
length and deviation distance from the centre of the bench model but still took
longer time to complete the task when compared with the proficiency goals.
This would infer that the hand motion of these trainees was economical and
efficient, similar to the expert surgeons but they were slower than their
superiors. However, time is a poor measurement of technical skills. The time
taken to complete the task could be improved by practice and experience in the
real setting. While operative speed is a desirable surgical quality to lower the
time spent under anaesthesia, it fails to assess the quality of surgical
performance (Shah et al., 2003).

We also used the Patriot™ to determine skills retention of the trainees while in
the training programme. In general, this cohort showed improvement in their
time taken to complete the task and their hand motion in both technical skills.
The difference in their skills performance was evident after a dedicated
bootcamp session at the beginning of their training, mandatory operative skills
classes in the simulation lab and five months of experience in the hospitals.

The automated Patriot™ system has a high potential in a learner-oriented
proficiency curriculum (Shaharan et al., 2015). By providing an objective and
numerical rating, trainees could benchmark and aim to improve their score
through enhancement of surgical skill (Saleh et al., 2009). When there is any
low or unsatisfactory performance detected, the training module can be
customized for the trainees to allow remedial session. During the remediation
process, the supervisor would highlight the positive and negative points in the
trainees' performance and any corrective action can be taught directly.
There has been myriad of surgical simulators reported in the literature. Since the year 2000 approximately 173 studies were published that specifically reported construct validity of a wide spectrum of surgical simulators (Shaharan and Neary, 2014). However, only a handful number of articles that described the use of these simulators in skills assessment in a surgical curricula and selection process. The simulators should be fully utilised in the simulation-based training structure. This is because of the advantage of being able to produce automated objective scoring system and does not require a group of observer to assess the performance.

Nevertheless, the remaining question is whether these skills in the simulation lab are transferrable to real operations. The fundamental assumption of simulation-based training is that the skills acquired in simulated settings are directly transferable to the operative setting (Sturm et al., 2008). According to a recent systematic review, Buckley et al (Buckley et al., 2013) demonstrated that simulation-based training has a positive impact on operative time and predefined performance scores in the operating room (OR) but not the quantifiable measures such as ergonomics, hand dominance and smoothness of movement as measured by simulators. The conversion from VR to OR as coined by Professor Anthony Gallagher (Seymour, 2008) has gained the attention of surgical educators recently and the application of simulators is beginning to take-off in the world of surgical training.
Chapter 6

Discussion & Future Work
6.1 The Aims of Thesis

The aim of the thesis was to identify an objective automated scoring system in order to assess fundamental basic surgical skills such as one-handed knot tying and suturing skills. It is important to properly assess these high stake skills in surgical trainees as failure to achieve proficiency can result in surgical catastrophes in the operating theatre. Therefore, an objective, reliable and valid assessment tool that does not rely on observers to analyse trainees’ performance would be most desirable in surgical training programme.

Previous literature showed a myriad of observational and non-observational tools to assess trainees performing simulated surgical procedures. These were mainly in minimally invasive surgical procedures such as laparoscopic skills. The majority of studies were targeted towards validating simulated non-observational tools to ensure that the devices are able to differentiate between different levels of experience and that the measurement corresponds to a previously validated assessment tool. This is usually the Objective Structured Assessment of Technical Skills (OSATS) system.

We analyse the strength and weakness of the existing OSATS in view of its validity and feasibility when used in assessing open basic surgical skills in our institution. In addition, there is paucity in the literature involving the assessment of open surgical skills using non-observational tools. This led us to analyse the validity and reliability of an off-the-shelf motion tracking device as an automated scoring system in open surgical skills assessment. We widened our research paradigm by demonstrating the potential use of this device in assessing our own surgical trainees without the expert observers.
6.2 Summary of Main Findings

- The Global Rating Scale (GRS) from the classic OSATS has to be modified by removing certain items when assessing open basic skills as the parameters were deemed irrelevant for the particular skill assessment.

- The modified GRS was able to differentiate different level of surgical experience in one-handed knot tying and simple interrupted suturing skills, hence construct validity was proven.

- We demonstrated that the components selected for the modified GRS measure the same construct in the basic skills assessment and therefore, it is a reliable observational tool.

- The inter-rater agreement or reliability of the modified GRS remain poor which indicates that the assessment was in some way subjective and the assessors could not come in to an agreement when assessing a trainee competency in fundamental surgical skills, especially in component “respect for tissue”. This identified the shortfall of observer dependent system.

- The automated motion tracking system or Patriot™ paired with our bespoke software was able to produce meaningful metrics in order to analyse hand motion when basic skills were being performed by users.

- The Patriot™ system was able to differentiate between the different levels of expertise (construct validity) and demonstrated significant correlation with the validated modified GRS (concurrent validity) in both one-handed knot tying and simple interrupted suturing skills.

- We demonstrated the potential application of the automated Patriot™ system in mapping trainees’ surgical proficiency by using the validated metrics. It can be used to monitor trainees’ progress throughout their training years and also to compare their scores with a pre-determined proficiency level.
6.3 General Conclusions

We have investigated in detail the validity and reliability of both observational and non-observational assessment tools for open surgical skills. We focused upon basic open surgical skills as there is limited availability of simulators or advanced devices that could monitor and assess the performance. This is in contrast to assessment of minimally invasive surgical skills. We discovered interesting findings on this subject.

We chose the classic observational tool, GRS from OSATS for the first part of this thesis. When we investigated the validity and reliability of the this tool, we eliminated components from GRS that was not relevant to the skills assessed in this study. We have proven the construct validity of the modified GRS in assessing basic surgical skills. This means that the scores from the modified GRS assessment could distinguish the performance from different level of surgical experience in these skills. In view of this finding, the modified GRS could be utilised to train and assess surgical trainees' skills in a proficiency-based training setting. When we analysed the internal consistency of the modified GRS, we found that all the components chosen has good correlation with each other which means there is a general agreement in between the components that form the structure of a valid modified version of this assessment tool.

Observational tool such as the GRS requires at least two observers or experts to assess the performance of the subjects or trainees in order to avoid any personal biases. We examined the level of agreement between the observers in the most ideal and independent set up. The inter-rater agreement was generally poor especially when the observers were assessing “respect for tissue”. The assessors could not agree on how good or bad the subjects were in handling the tissue while performing the tasks during majority of the assessment times.

This skill is an especially important part of any operation and should critically be assessed accurately. In addition to this disadvantage of this observational tool, it is labour intensive and requires a significant amount of time spent to assess
the subjects or video recordings. We certainly found that it was challenging to prepare the video recordings including labelling them anonymously and assess each one of them during an independent setting for this study.

In order to eliminate the need for expert observers or assessors, we explored the use of an automated motion tracking system in open surgical skills. Motion analyses have been widely used in other non-surgical areas such as in training spray painters, video games, ultrasound simulation and sports training. Its use in surgical training is, however, still limited to minimally invasive skills (MIS) such as laparoscopic skills by using simulators. The MIS simulators are able to track the movement of the instruments in a closed box and analysed them into objective scores. Lord Ara Darzi and his team developed a motion tracking system called Imperial College Surgical Assessment Device (ICSAD) to assess open surgical skills over a decade ago. The ICSAD analysed the hand motion and produced objective scores including total path length, time and number of hand movement.

We employed the same principle but using an off-the-shelf electromagnetic device called the Patriot™ and we created our own bespoke software to interpret the raw data from this device. The software was able to generate similar scores as ICSAD which were total path length and time. It also generated further scoring metrics which were average deviation distance from the centre of bench models and x-, y- and z-axis. In practice, the experienced surgeons or trainees have their hands moving at certain distance from their working areas which would not be too close or too near. This would ensure their hands are working in the most efficient and economical way. We aimed to determine this ideal working distance in basic surgical skills by using this motion tracking device.

We analysed the validity of the Patriot™ hand motion tracking system rigorously, similar to assessment of the GRS. Construct validity was demonstrated in time, path length, average deviation distance from centre of bench models and Z-axis for one-handed knot tying skill. When the Patriot™ system was used to assess simple interrupted suturing skills, construct validity was demonstrated in time, path length and average deviation from Z-axis. This
system was able to obtain a quantitative measurement of the hand motion and differentiate them according to the level of experience.

Another invaluable feature of this novel system is that the metrics showed a consistent correlation with the observational tool or GRS and therefore proved the concurrent validity. The main finding was that the metrics from the Patriot™ system correlated well with the GRS items, specifically with components assessing motion and flow of operation. We could safely suggest that this automated system has a great potential in replacing the observational tool in assessing basic surgical skills and therefore reduces the need to have observers for the assessment session.

According to our literature review, there is a paucity of published study in using the non-observational tool to assess surgical trainees following extensive validation of these tools. In Chapter 5, we explored the application of the Patriot™ system in the Irish surgical training programme. We applied this system in a proficiency-based assessment. A proficiency level was set by a group of surgical consultants. We then mapped out the trainees’ scores and we were able to identify trainees who did and did not perform well in the open skills.

In addition to this, we performed an assessment of fundamental skills among trainees before and five months after commencing surgical training programme in our institution. We were able to objectively report their progress using the Patriot™ system. The need for observers to assess them was eliminated because the system is automated and provides the score instantly. The scores were objective and numerical which makes it easier for trainees or supervisors to analyse the performance over a period of time in surgical training. Any trainees with low performance can be identified earlier and extra practices would be required in order to achieve proficiency. This can be tailored to each trainee’s needs.

Overall, this thesis has demonstrated that the observational tool is subjective, although it is a valid tool to assess open basic surgical skills. It has established the validity of an automated hand motion tracking system to assess the fluidity of the hands when completing these fundamental skills. This thesis has taken a step forward beyond the validation paradigm of an automated hand motion tracking system. It has demonstrated the potential application of this powerful
system in the surgical training programme. We strongly suggest that it should be utilised in the assessment of surgical trainees’ technical skills and also be a part of the selection process in order to select the best candidates for future surgeons.

The limitation of this thesis is the relatively small sample size used in the experiments. We decided to use a small sample size due to the practicality of the experiment sessions. The experiments were time consuming and labour intensive. The subjects were recruited during the mandatory teaching sessions such as Human Factor or Operative Surgical Skills classes in the Royal College of Surgeons in Ireland. Since the participation was voluntary, surgical trainees were required to perform a set number of tasks either during their breaks or after the teaching sessions. The recruitment of consultant surgeons as experts was challenging too. They were invited to perform the skills before or after the teaching sessions and the number of surgeons who were involved in these classes were limited. We could not recruit more subjects in the hospital settings due to logistic difficulty in setting up the experiments outside the research facility and it would disturb the integrity of the experiment’s standardisation.

Another limitation of this thesis is that we did not analyse the outcome of each tasks. As surgical competence is multimodal, we only analysed the hand motion and global performance in the experiments. The quality of the end products after completing each task is yet to be analysed. This is part of an ongoing and future work.

All the experiments were performed using bench models in a simulation laboratory. We did not establish any correlation between the performances in this setting with performances in a real operating theatre situation. Therefore the transferability of our results to the real life operation is uncertain. As a result, we could not draw a conclusion that the evidence from simulator based performance resembles the performance in real life settings.

Finally, we did not assess the performance of subjects on a more complex skills or a complete procedural task. We only assessed basic surgical skills which are the most fundamental skills to grasp at an early stage of training before being able to perform more complicated task or a complete procedure. In addition, it is time consuming to record and assess complete procedural tasks compared to
basic tasks. There is no doubt that it is important to demonstrate the applicability of the motion analysis in a setting of a full complete procedure in open surgery. We have undertaken an experiment to assess the validity of the motion analysis in a simple complete procedure task which is excision of sebaceous cyst. The preliminary result is promising and we anticipate that the use of motion analysis in surgery has no defined boundaries.

6.4 Future Works

The research presented in this thesis has raised many questions and should be pursued in future potential studies. First of all, it is worth considering the assessment of the quality or outcome of a specific surgical task. We propose that the end products should be analysed along with the motion analysis. This innovation would create an all-in-one package in assessing surgical competency in all training bodies worldwide.

Another proposal for future research is an evaluation of the motion analysis system use in a real operating theatre. Currently, the motion analysis system is sensitive to surrounding metal objects which can cause erratic reading. New invention of a system that can be used in real operating theatre is much desired.

Finally, a further study in assessing the correlation between motion analysis outcomes in the simulation laboratory and in real operating theatre settings should be pursued in the future. The result from this work would change the way we assess surgical trainees in order to be able to produce a group of top quality surgeons.
Appendices
Appendix I  Written Instruction for Each Task

Task 1: Knot Tying, 1-handed, no tension

- Instruction: Tie a 46cm 2/0 silk ligature around 5mm white segment on the rubber tubing using 1 handed technique, **3 square knots with 6 throws**
- Cutoff Time: 60s

Task 2: Suturing, Simple, Interrupted

- Instruction: 30cm suture, pass needle through 2 inked targets, instrument tie surgeon’s knots and 2 square knots (3 knots in total)
- Cutoff Time: 120s
Appendix II  Letter of Ethical Approval

Royal College of Surgeons in Ireland
The Research Ethics Committee  
121 St. Stephens Green, Dublin 2, Ireland.  
Tel: +353 1 4022373  Fax: +353 1 4022205  Email: recadmin@rcsi.ie

Dr. David Smith, Acting Chair  
Dr. Niamh Clarke, Convener

4th July 2013

Ms Shazrinizam Shaharan  
Department of Surgical Affairs,  
121 St Stephen’s Green,  
Dublin 2

<table>
<thead>
<tr>
<th>Ethics Reference No:</th>
<th>REC775b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project Title:</td>
<td>Design of Objective Scoring System for Basic Surgical Skills (BSS)</td>
</tr>
<tr>
<td>Researchers Name:</td>
<td>Shazrinizam Shaharan</td>
</tr>
<tr>
<td>Other Individuals Involved:</td>
<td>Mr Donncha Ryan (Learning Development Manager), Department of Surgical Affairs, RCSI</td>
</tr>
</tbody>
</table>

Dear Shazrinizam,

Thank you for your Research Ethics Committee (REC) application. We are pleased to advise that ethical approval has been granted by the committee for this study.

This letter provides approval for data collection for the time requested in your application and for an additional 6 months. This is to allow for any unexpected delays in proceeding with data collection. Therefore this research ethics approval will expire on 4th March 2014.

Where data collection is necessary beyond this point, approval for an extension must be sought from the Research Ethics Committee.

This ethical approval is given on the understanding that:
- All personnel listed in the approved application have read, understand and are thoroughly familiar with all aspects of the study.
- Any significant change which occurs in connection with this study and/or which may alter its ethical consideration must be reported immediately to the REC, and an ethical amendment submitted where appropriate.
- Please submit a final report to the REC upon completion of your project.

We wish you all the best with your research.

Yours sincerely,

[Signature]

PP Dr. Niamh Clarke (Convener)  
Dr David Smith (Acting Chair)
Subject Information Sheet

“The Design of an Objective Scoring System for Basic Surgical Skills (BSS)”

Principal Investigator: Mr Paul Neary
Co-Investigator: Ms Shazrinizam Shaharan (Shaz)

You are being invited to take part in a research study. You should read the information below carefully and take your time to answer the questions - please do not feel rushed or under pressure.

What the study is about.

This study is about identifying metrics for assessing basic surgical skills (BSS) objectively using the tracking tools. We would like to determine the pattern of learning to proficiency in surgical skills among surgical trainees. This, in turn will create a scoring system of BSS for the trainees by using objective measurement and integrate it into the BST curriculum.

Your participation in this study is voluntary.

What you will be asked to do.

You will first be asked to sign a consent form if you agree to take part. Then, you will be asked to complete a questionnaire which contains simple questions in regards to your personal details and previous surgical experience. The next step, you will be asked to perform a set of open surgical skills by using basic surgical set and the tracking device. The tracking device will track your hand movement while performing the tasks and the data will be analysed at a later stage.

Confidentiality

All the information and results from the tests will not be released to or shared with any surgical or medical training bodies and will have no effect on your career. All the data will be stored in a secure location.

Implication

By taking part in this study, it will give you the chance to familiarise and practice on open surgical tasks such as knot tying, suturing and making a laparotomy incision. It will expose you to a new technology that could potentially be used in surgical skills assessment in the future. There are no known anticipated risks to you by participating in this study. The tracking device is a non-invasive tool and should not cause any discomfort or pain during or after using it.

Contact Details

If you have any questions, concerns or complaints about this study, you can talk to the consultant who is coordinating the study or to a member of the study management team at RCSI.

RCSI contact: Ms Shazrinizam Shaharan
Email: shazrinizamshaharan@rcsi.ie
Phone No.: 01-4022704

Supervisor: Mr Paul Neary
Phone No.: 01-4142000
Appendix IV Consent Form for Participants

Consent Form
Form of consent by a subject volunteering to take part in research associated with surgical training

Consent By the Subject
I ________________________________ (full name)
of ________________________________ (address)
hereby fully and freely agree to take part in a study entitled

“The Design of an Objective Scoring System for Basic Surgical Skills (BSS)”

I have read and understand the subject information sheet.

I understand that any data collected during the trial which pertains to me will be treated in a confidential manner.

I understand that this study is designed to promote scientific knowledge, develop surgical training and improve patient care.

I agree to fulfil any training requirements to the best of my ability.

I understand that I may withdraw my consent at any stage in the study without having to provide an explanation. I acknowledge the purpose of the study. The nature and purpose of any procedure has been described and explained to me by Ms Shazrinizam Shaharan and I have discussed these matters with her to my satisfaction.

I hereby give my consent for photography, filming, videotaping and/or audio recording or other means to capture my image or voice and/or being quoted in printed materials.

Signed: __________________________

Print Name: ______________________

Date: __________________________
REFERENCES


LIST OF PUBLICATIONS AND PRESENTATIONS

Publications


In Press


Oral Presentations

SHAHARAN, S., SEKHON, R., RYAN, D.M., TRAYNOR, O., NEARY, P. Objective measurement of knot tying skills using motion tracking system” Presented at Sir Peter Freyer Surgical Symposium, NUI Galway, Ireland. Sept 2013.

SHAHARAN, S., RYAN, D.M., TRAYNOR, O., NEARY, P. Basic surgical skill retention: Can Patriot™ motion tracking system provide an objective measurement for it? Presented as short paper at the International Surgical Congress of the Association of Surgeons of Great Britain and Ireland (ASGBI), Harrogate, UK, April 2014.

SHAHARAN, S., RYAN, D.M., TRAYNOR, O., BUCKLEY, D., NEARY, P. **Global Rating Scale (GRS) under the microscope.** *Presented at Sir Peter Freyer Surgical Symposium, NUI Galway, Ireland, Sept 2014.*


**Poster Presentations**