Abstract

“Tell me and I forget, teach me and I may remember, involve me and I learn”,

Xun Kuang

The surgical training environment has been required to adapt to multiple changes in recent years. Two challenges have placed significant burdens on traditional training methodology: the emerging need for critically reflective clinicians, and restrictions in the traditional methods of technical skills training. Implementing training for these disparate elements is challenging. Simulation-based training has been widely used to teach both technical and non-technical skills to medical students and residents, however consistent access to faculty feedback during training and implementation of critically reflective activities remains problematic. Self-regulated learning can be used to both provide the mechanism by which skill acquisition and critical reflection can occur. This research consisted of two complimentary studies examining technical skill acquisition and self-regulated learning in medical students using structured peer-feedback to learn technical skills.

The technical skills study consisted of two prospective randomized controlled trials comparing structured peer-feedback versus traditional faculty feedback for the acquisition of a technical skill. Participants in the peer-feedback group provided feedback to each other using a structural process taught through pre-training with an assessment tool. Participants completed five attempts at the skill over four days. Performance in the peer-feedback group was found to be non-inferior to the faculty group across all assessment criteria used in the study: blinded expert assessment, time to completion and functional test.
The impact on self-regulated learning was examined using thematic analysis of student responses to multiple surveys and the written feedback provided by the peer-feedback group. Students developed increasingly sophisticated assessment and feedback skills through regular small exposures to the assessment task. Evidence of self-regulated learning was demonstrated by students’ improvement in technical performance, evaluative judgement of their peer’s performance, self-identification of the benefit of these assessment skills to their own learning, and subsequent implementation of these skills in other contexts. Despite improved assessment skills students did not believe the feedback they provided was beneficial – likely due to a lack of external feedback on their assessment skills from faculty.

This research has demonstrated that with appropriate support, structured peer-feedback can be used to augment traditional faculty led training, whilst additionally developing students’ evaluative judgement and self-regulated learning.
Co-Authorship

Dr Richard Reznick (Co-primary Supervisor and thesis committee member) was involved in the study design, analysis of the studies and the editing of the manuscripts and thesis preparation.

Dr Boris Zevin (Co-primary Supervisor and thesis committee member) was involved in the study design, execution, analysis of the studies as well as the editing of the manuscripts and thesis preparation.

Dr Don Klinger (Faculty of Education Supervisor and thesis committee member) was involved in the development of the research question, execution of the studies, manuscript editing and thesis preparation.

Dr Leslie Flynn (thesis committee member) was involved in the guidance of the studies, manuscript editing and thesis preparation.
Acknowledgements

I would like to thank all those who have made this unexpected, but unmissable journey possible. To Richard I want to thank you for your encouragement to think bigger, push harder and aim higher than I thought possible, both regarding this Master’s, but also more importantly, the future. To sit under your wing for this was life changing and it was a great privilege to be able to be your fellow. To witness your example as a medical educator was incredibly encouraging, as it showed me what is possible, with hard work, great mentors, a supportive family, and a little providence.

To Boris I owe a great debt of gratitude. From those initial discussions about research options whilst sitting on my in-law’s veranda in warm Australia, through to the concepts of statistical analysis and research design, I’m exceedingly grateful. I have been encouraged by watching how an emerging medical educator struggles to find space to meet all those competing demands yet remain humble.

To the other members of my committee, Don, thank you for your support and guidance throughout the Master’s. Guiding me through the education system was incredibly helpful, as was your wisdom with the design and development of the research. Your kind words about my work has given me cause to pause and consider what the future might hold. Leslie, thank you for your gently challenging way that helped me identify and articulate what I was really trying to research.

To the other faculty in Education, thank you for your openness, willingness to encourage me in this area of learning. John, Rena and Azza, your classes have inspired what is contained within this thesis, and has made lasting impressions on what research I am interested in.
The Queen’s Medical Simulation centre team, Kim, Loren and Jeremy, thank you’re your generosity, insights and experience that allowed me to complete this research in your facility.

To the Royal Australasian College of Surgeons and the individuals involved in the scholarship scheme, thank you, for your support through the Fellowship of Surgical Education scholarship, without which this journey would have never occurred.

To my wife Rebecca and daughter Hannah, without your willingness (begrudgingly given at times) to be involved in this crazy adventure, this Master’s would not have been possible. This adventure would not have been any fun without you, and I’m glad that in the end, this journey has made us stronger than we were, as individuals and as a family. Your love, support, encouragement and help has been instrumental in the completion of the Master’s, so thank you. I love you both.
Table of Contents

Abstract ................................................................................................................................................... ii
Co-Authorship ........................................................................................................................................ iv
Acknowledgements ................................................................................................................................. v
List of Figures ........................................................................................................................................ ix
List of Tables .......................................................................................................................................... x
List of Abbreviations .............................................................................................................................. xi
Chapter 1 Introduction ............................................................................................................................. 1
  Current Challenges to Teaching and Learning Technical Skills ............................................................ 1
  Current Challenges to Developing Independent Life-Long Learning Skills ........................................... 2
  Research Questions ............................................................................................................................. 3
  Hypotheses .......................................................................................................................................... 3
Chapter 2 Literature Review .................................................................................................................... 4
  Introduction ......................................................................................................................................... 4
  Well defined task: how assessment can both characterize and drive performance ................................. 6
  Highly motivated individuals: how personal motivation influences performance .............................. 14
  Repetition: facilitating repetition in medical education through simulation-based training .................. 18
  Feedback: using feedback to improve performance and assist learners in building their self-regulated
    learning skills .................................................................................................................................... 23
Chapter 3 Technical Skills Study ........................................................................................................... 33
  Abstract ............................................................................................................................................. 33
  Introduction ....................................................................................................................................... 34
  Methods ............................................................................................................................................ 35
  Results .............................................................................................................................................. 41
  Discussion ......................................................................................................................................... 50
  Conclusion ........................................................................................................................................ 55
Chapter 4 Self-Regulated Learning Study .............................................................................................. 56
  Abstract ............................................................................................................................................. 56
  Introduction ....................................................................................................................................... 58
  Results .............................................................................................................................................. 63
  Discussion ......................................................................................................................................... 70
Chapter 5 General Discussion ................................................................................................................ 78
  Technical skills study ........................................................................................................................ 78
List of Figures

Figure 1. Hattie and Temperley's model of effective feedback............................................................... 24
Figure 2. Anastomotic leak test. .............................................................................................................. 38
Figure 3. Skin excision and closure stress test. ....................................................................................... 39
Figure 4. Bowel anastomosis: expert OSATS scores by attempt ............................................................ 43
Figure 5. Bowel anastomosis: completion time by attempt ...................................................................... 44
Figure 6. Bowel anastomosis: anastomosis leak pressure by attempt ..................................................... 45
Figure 7. Skin excision: expert OSATS scores by attempt ................................................................. 47
Figure 8. Skin excision: completion time by attempt ............................................................................. 48
Figure 9. Skin excision: stress test by attempt ....................................................................................... 49
List of Tables

Table 1. Bowel anastomosis: performance at baseline and fifth attempt................................................. 42
Table 2. Bowel anastomosis: OSATS scores subdivided into checklist and global rating scale. ............... 43
Table 3. Skin excision: performance at baseline and fifth attempt. ............................................................ 46
Table 4. Skin excision: OSATS scores subdivided into checklist and global rating scale. ....................... 47
Table 5. Responses for Likert survey questions. ..................................................................................... 50
Table 6. Frequency of theme 1: non-specific OSATS feedback .............................................................. 64
Table 7. Frequency of theme 2: quality-related feedback........................................................................... 65
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>B</td>
<td>Bowel Anastomosis</td>
</tr>
<tr>
<td>DP</td>
<td>Deliberate Practice</td>
</tr>
<tr>
<td>EVT</td>
<td>Expectancy Value Theory</td>
</tr>
<tr>
<td>FF</td>
<td>Faculty Feedback</td>
</tr>
<tr>
<td>GRS</td>
<td>Global Rating Scale</td>
</tr>
<tr>
<td>OSATS</td>
<td>Objective Structured Assessment of Technical Skills</td>
</tr>
<tr>
<td>PF</td>
<td>Peer Feedback</td>
</tr>
<tr>
<td>S</td>
<td>Skin excision and suture closure</td>
</tr>
<tr>
<td>SDT</td>
<td>Self-Determination Theory</td>
</tr>
</tbody>
</table>
Chapter 1
Introduction

Current Challenges to Teaching and Learning Technical Skills

The postgraduate education environment has undergone several significant shifts in the past ten years. There have been many forces that have resulted in an effective decrease in the amount of time and independence future trainees are getting in the operating room to learn the technical aspects of surgery. Introduction of safe working hours, patient safety concerns, increasing numbers of trainees and use of Clinical Fellows on units, and economic and efficiency drives from hospitals have all affected the accessibility of learning technical skills.

Competency based medical education rather than time-based apprenticeships has been developed to respond to some of these demands; however, even this approach to teach still requires planned incremental access to technical skills training.

Simulation has helped fill the gaps, but this requires the availability of: equipment, disposables, educational curriculum and an expert to provide feedback. Expert feedback has been found to be more effective than self-assessment in improving technical performance (Porte, Xeroulis, Reznick, & Dubrowski, 2007). Yet it is not clear if the presence of an expert is essential during the practice phase of learning. As one example, recent studies showed that participants who compared their own recorded performance with an expert video had similar performance improvements compared with participants reviewing their performance with an expert present (Nesbitt, Phillips, Searle, & Stansby, 2015).
Current Challenges to Developing Independent Life-Long Learning Skills

Theorists have explored how successful learners combine self-motivated beliefs and behavioral learning strategies in a self-regulated way to achieve their goals. This deliberate process occurs in 3 stages: a metacognitive forethought phase, an intra-performance control phase, and a post event self-reflective phase (Zimmerman, 2002). Self-directed learning and other self-regulated learning styles (Zimmerman, 2002) are associated with higher academic performance (Nicol, 2009; Tagawa, 2008), deeper understanding of topics (Räisänen, Postareff, & Lindblom-Ylänne, 2016), higher academic efficiency in Doctoral students (Kelley & Salisbury-Glennon, 2016) and greater life-long learning (van de Wiel & Van den Bossche, 2013; van de Wiel, Van den Bossche, Janssen, & Jossberger, 2011). Although studies of medical trainees have confirmed the utility of this theoretical framework in academic contexts, it is perhaps the lack of focus on the self-reflective phase, with its own inherent skillset that may be causative of the poor accuracy in self-assessment found by Davis et al. (2006) and Savoldelli et al. (2006). Providing medical learners with the skills necessary to perform this evaluation would likely improve the learners self-regulated learning in addition to improving their professional and clinical acumen. The term evaluative judgement will be used in this thesis to describe this evaluation process and the skills used to perform it. Unfortunately, the terms self-reflection or self-assessment, although describing these evaluation processes, are highly value-laden and therefore can be confusing in today’s education literature (Desjarlais & Smith, 2011).

Studies of technical skill acquisition have thus far only explored expert feedback or incidental unscripted self-assessment and feedback. Yet no study has looked at teaching students how to critically evaluate their performance, and the subsequent impact this has on their performance and their self-regulated learning.
Research Questions

1. How does technical skill acquisition compare using structured peer-feedback to traditional faculty feedback?

2. How do the participants’ peer-feedback skills develop during acquisition of technical skills?

3. How does the teaching and practice of peer-based assessment and feedback during acquisition of technical skills impact students self-regulated learning?

Hypotheses

1. In learning a technical skill in a simulated environment, structured peer-feedback via video, is not inferior to traditional faculty feedback.

2. The quality of the participants’ peer-assessment and feedback skills will increase as they acquire technical skills.

3. Peer evaluation and feedback skills can be integrated into participants’ self-regulated learning armamentarium, when appropriately integrated into a technical skills study.

This thesis uses a manuscript-based format. As such each chapter consists of a standalone manuscript addressing one of the research questions. Preceding these chapters is a literature review that provides background to the research and augments the literature discussed within the manuscript.
Chapter 2

Literature Review

Introduction

The approach to training medical doctors has undergone significant changes over the past sixty years. Currently, we stand at a new crossroad for medical education. New societal pressures such as safe working hours, patient safety concerns, increasing numbers of trainees, and economic and efficiency drives from hospitals and health departments have all affected the accessibility of surgical and other residents to traditional methods of learning technical skills (Bolster & Rourke, 2015; “Competence by design,” 2014; Temple, 2014). This confluence of different pressures creates the potential that qualified doctors do not have the skills or capabilities that align with societal expectations. In the traditional apprenticeship model, used for the last 100 years, trainees were taught skills by qualified physicians and through close supervision were given graduated amounts of independence (Blinman, 2017). This resulted in an opportunistic approach to learning which had the potential for gaps to develop in a trainee’s experience (Knox, Gilardino, Kasten, Warren, & Anastakis, 2014).

One method of mitigating these potential gaps is by using simulation-based training (Ziv, Wolpe, Small, & Glick, 2006). Simulation-based training defined as a requisite skill or process that is simulated in an educational environment, allowing educators and trainees to focus specifically on learning, without risk of harm to others (Jones, Passos-Neto, & Braghiroli, 2015; Levine, DeMaria, Schwartz, & Sim, 2013). Medical simulation-based training entered clinical education practice in the 1960’s. In the United States, the Sim One patient simulator was developed (Levine et al., 2013), and in Europe, the Resusci-Anne® cardiopulmonary resuscitation trainer was being used (“Resusci Anne QCPR,” 2018). Aside from technical skills
mannequins, other simulation-based training modalities such as standardized patients were also being developed during this time. Although simulation-based training remains heavily grounded in technical skills training, increasingly it is being used for non-technical skills such as communication and teamwork. As with the airline industry, these essential, non-technical skills are recognized as major contributors to poor patient outcomes (Levine et al., 2013), and so opportunities to practice and develop these skills in a safe environment promote better patient outcomes.

Most novices can obtain a degree of automatic competence for simple tasks within approximately 50 hours of practice (Ericsson, 2008), but to move to expert performance further element is required. Based on his research with chess masters and musical virtuosos, Ericsson was able to replace the traditional view: that of experts having some innate advantage (Galton, 1892), with the concept of deliberate practice (DP). Experts used an intentional or deliberate practice regime that allowed them to reach higher levels of performance. Deliberate practice incorporates four key concepts: 1) a well-defined task that facilitates the assessment of a performance at the highest levels; 2) a high level of motivation by the learner to improve their own performance; 3) immediate relevant feedback; and 4) ample opportunity to practice specific components and refine existing performance. The effectiveness of this approach to skill development in the medical environment has been subsequently confirmed in multiple studies and meta-analysis (e.g., McGaghie, Issenberg, Cohen, Barsuk, & Wayne, 2011).

Each section of this literature review examines one of the four elements of deliberate practice as it pertains to the literature on technical skill acquisition in a medical education context. Questions and gaps that arise in the literature will be identified to guide the development of the research questions addressed by this thesis.
Well defined task – how assessment can both characterize and drive performance

The first principle of deliberate practice is a well-defined task. Defining a task has dual purposes. It firstly allows a complex task to be reduced into its component parts using reductionist techniques like metrics or rules. For example, swimming 100m butterfly is defined by certain arm and leg movements, a specified method of start and finish, and a 100m distance. This task is definably different from swimming 100m freestyle. These rules or categorizing information, in addition to merely defining what a 100m butterfly swim is, can also be used for assessment purposes. Assessment can be applied to the whole task, who won a 100m butterfly race by staying within the rules, as well as within the task, who has the best start, or the best stroke, etc. It is this secondary assessment purpose for which Ericsson sought to accurately define a task. Experts identify component areas of weakness to be strengthened using practice. This allows learners to achieve their high level of performance, rather than rehearse the whole task or areas of strength. This section will examine common methods of assessing performance of a surgical task, focusing primarily on expert-rater tools, and explicitly on the tool chosen for my research. The mechanisms for providing evidence of validity of score obtained using these assessment tools are discussed, before finally examining alternative methods of technical assessment.

In medical education, experienced physicians and surgeons in particular have long been able to intuitively differentiate between expert and novice performance on procedural tasks, however articulating this difference in performance has been more difficult, resulting in years of research (e.g., Aggarwal et al., 2010; Martin et al., 1997). The results of this research have allowed educators to both clearly define a surgical task and subsequently make judgements regarding a performance. Surgery like other performance-based tasks such as music, or dance,
requires the assessment to be conducted during the performance itself, as there is rarely an artefact to examine afterwards. To conduct a technical performance assessment requires: a measurable construct, such as efficiency, a means of observing that construct, and a standard to compare the performance against.

Fundamentally, tools developed to assess technical performance can be categorized into two types, based on the construct and the method of observation. In common with all these assessment tools is the need to compare a performance to a referenced standard – which is often chosen to be an experts, but not always (Linn & Miller, 2005). The first group of tools are those that use assessor-based inferences anchored on different elements about the performance. Trained raters observe the task live or on video and score the performance based on the defined anchors that consist of a mixture of agreed upon checklist items or global rating scales. Examples of this method include Objective Structure Assessment of Technical Skills, Global Assessment of Laparoscopic Surgery, and Operative Performance Rating System (Ghaderi et al., 2015). The second and smaller group are those that use technology to measure a physical characteristic of the performance such as, pressure applied by the instrument, or number of movements to complete a particular task (Porte et al., 2007).

Checklists reduce a task into a number of activities or components that must be performed during the procedure, usually in chronological fashion. This allows the assessor to determine if the procedure is performed in the prescribed manner. There are two limitations with this method. The first is that yes/no checklists are primarily focused only whether steps were completed correctly, additional checklist items must also be added to assess quality or efficiency of the overall performance. The second limitation is the prescriptive nature of the component steps that have been used to create the checklist. Variations in procedural techniques are
universal, so if the checklist is too rigid, then alternative techniques will be penalized. This inflexibility can be avoided by seeking advice from multiple experts about the requisite components in the checklist (Zevin et al., 2013). Experts are invited to develop or discuss the necessity of each item potentially contained within a checklist. Once agreement of each item has been reached – usually 80%, then the item is included within the checklist. This method of obtaining consensus about a topic is known as the Delphi method (Palter & Grantcharov, 2012; Zevin et al., 2013).

The second approach commonly used in assessor-based tools is global rating scales (GRS). GRS rate specific characteristics on a defined scale of efficiency of performance, typically 1 to 5, that contain descriptive anchors that define each level of performance. For example, time and motion of the procedure: 1 (many unnecessary moves), and 5 (clear economy of movement and maximum efficiency). This method is the complement of the checklist, assessing the quality of the performance. It also divides the task into composite themes, such as instrument handling, tissue handling and knowledge of the procedure, to provide a structured assessment of the overall performance, rather than the more reductionist approach of checklists. This unfortunately leads to one of the main limitations of GRS as GRS provide an assessment of the overall performance. For example, a GRS score of 3 is unable to provide participants with meaningful feedback to assist learners to improve their performance. This can be overcome with additional written comments justifying the reasons for such a score. The objectivity of the GRS is also dependent on the specificity of the written description of the performance for each anchor, and the assessor attitude to scoring. It has been suggested that improved interrater reliability between assessors can be achieved by assessor training (Ghaderi et al., 2015). Given that
checklists and GRS act in a complementary fashion, it is not surprising that most performance assessments use a combination of the two (Levine et al., 2013).

The assessor based assessment tool I have chosen to use for my thesis research is the objective structured assessment of technical skills (OSATS) (Martin et al., 1997). The OSATS consists of a procedure specific checklist and GRS that are completed by trained evaluators during a live or video recorded performance (Dath et al., 2004). Originally developed for a benchtop model examination, the OSATS tool has been translated into the operating room for a number of different procedures, both open and laparoscopic. OSATS, was one of the first procedural assessment tools developed prior to the advent of the unitary validity construct, so most of its evidence was developed to meet the previous requirements. Given that the traditional validation construct focused on the test and on the interpretation, the subsequent literature involving OSATS has varying levels of evidence of context specific validation of the OSATS scores itself. Newer assessment scales such as Fundamentals of Laparoscopic Surgery, or Global Objective Assessment of Laparoscopic Surgery may have stronger levels of evidence (Ghaderi et al., 2015) but cannot be translated to open surgical procedures.

Theorists (Messick, 1995) recognized that the educational goal of assessment is not to create a perfect test, but rather to use the information gained from the test to inform decisions about learning. Evidence must be gained from the social environment or context in which this assessment occurs as part of the validation process for example, a high stakes credentialing examination, versus a national test for assessing numeracy and literacy skills in children. The outcome of both tests may have significant consequences for the individual and society, which assessment developers should be cognizant of. This assessment of consequences and context remains controversial with Mehrens (1997) arguing that it is beyond the remit of assessment
developers to determine the consequences of the assessment, particularly when misinterpretation is done for political purposes. Very few studies in the medical procedural assessment area explore the issue of consequence (Ghaderi et al., 2015), most of the articles identified by Ghaderi et al. (2015) use the assessment for feedback purposes only – formative assessment. Publications defining competency thresholds for summative assessment have been limited (Szasz et al., 2017).

Messick’s unitary validity framework has been adopted as the standard by which validity of assessment tools are assessed (Ghaderi et al., 2015). This framework recognises that there are limitations with each tool, but of greater importance is how the tool fits with the other sources of evidence thereby assisting educators make decisions about learner’s achievement. The unitary framework incorporates previous validity evidence frameworks (content, construct, criterion), but now seek evidences from five major sources: content, response process, internal structure, relationship to other variables and consequences. How evidence can be acquired for each source is briefly described below.

Content evidence is ensuring that the content assessed within the tool is relevant to the intended measurement. This is often achieved by using expert opinion, via such means as a Delphi method to ensure appropriate content is included. Response process evidence examines quality control elements such as through using blinding of raters, rater training and scoring analyses. Internal structure evidence examines the reliability of the scores across test elements, such as inter-item reliability, question factor analysis, scoring correlations etc. Relationship to other variables examines how this tool aligns with other performance measures, often labelled the “gold standard”. Finally, consequence evidence seeks to gather evidence of the consequence
of the results from using the scores from the tool. This often involves setting pass/fail marks and the impact of these scores on learners, e.g. failure to progress.

Based on this acceptance of Messick’s framework, multiple systematic reviews on technical skills assessments have been completed (Borgersen et al., 2018; Cook, Brydges, Zendejas, Hamstra, & Hatala, 2013; Ghaderi et al., 2015). Common to all is the lament that many technical skills assessment tools are unable to address all the elements of the unitary framework, commonly due to lack of developer familiarity with this unitary framework, or lack of consequence evidence being provided. As stated by Borgersen et al. (2018) “On the basis of the results of this systematic review, we would like to emphasize the need for exhaustive and methodologically sound investigations of validity evidence for assessments tools used in surgical specialties”. This need for exhaustive investigations highlights some of the limitations associated with using this unitary approach, or rather the use of published data for gathering such information.

The time required to obtain exhaustive evidence can be significant, as evidenced by the time taken to provide validity evidence for Fundamentals of Laparoscopic Surgery (FLS) examination (Peters et al., 2004). High-stakes assessments such as FLS, implemented by National bodies, require extensive validation, but this research occurs outside of regular training curricula by education experts. This in contrast to small programs or faculty seeking an immediately available formative assessment tool for an identified training gap. Herein lies part of the problem with requiring such extensive validation prior to the assessment tools implementation, no program or faculty can afford to halt training whilst waiting for the validation to be completed before implementing the assessment tool. This either means that tools are implemented without adequate validation or well validated, but perhaps inadequate tools are
used, merely because of their acceptability. For example, from Ghaderi et al.’s (2015) review only five tools met Messick’s framework, does this mean that these tools are the only ones that should be used? Or as Ghaderi comments, should we “focus on improving existing tools rather than re-inventing new ones”. Available tools may have adequate evidence for use in formative assessment, but if such tools are to be used for in a more summative way, such as to limit progression of skills outside of the simulation laboratory, is there enough validity evidence for the tool to be used in that way, or is evidence only obtained after the tool is implemented in this way?

A further limitation highlighted by Sweet et al., (2010) is that the unitary approach recognises that these assessment tools occur within a curriculum, and as such it should not just be the tool, but rather the whole educational package that should be included within the validation process. Although consequence evidence such as marking schemes, is required within the unitary framework, little of this broader view on the validation of the whole training curriculum has been discussed by the recent meta-analyses highlighted above. Furthermore, the need for consequence validity remains debatable, with Mehrens (1997) arguing that it is beyond the remit of assessment developers to determine the consequences of the assessment, particularly when misinterpretation is done for political purposes.

These limitations highlight that validation of technical tools using Messick’s unitary validity framework is more difficult than it appears. Despite these limitations, Messick’s unitary approach is the standard against which assessment tools are measured, and so validation must be described using this framework. Messick’s framework recognises that no tool is perfect, and so knowledge of this should guide faculty into seeking out the tool most appropriate for their needs.
The second method of assessing technical performance was through the use of technology to observe a physical characteristic of performance. One thrust of this research has been in using motion capture software to examine the movement characteristics of performance and compare those to expert performance (Vivek, Chang, Mackay, & Darzi, 2002). The rationale behind this is that experts are more economical in their movements, they are more fluid in the movements they make, and are more consistent with their tools, for example, able to apply the same amount of pressure on a tissue repeatedly. This method of assessment is highly suited to computer simulators such as those that use virtual reality (Nickel et al., 2015; Porte et al., 2007). Although highly suitable for differentiating between novice and expert performance, such information is difficult to interpret in formative settings because similar to global rating scales, it is the interpretation of these criteria that helps learners define why performance is lacking and mechanisms that can be used to improve upon it. For example, in research on student self-assessment, only the number of hand movements compared with experts was understandable to students as they self-assessed, despite researchers (Porte et al., 2007) being able to collect data on a variety of other criteria.

The last alternative method of assessment is examining the physical properties of the final product. This is not always possible in procedural skills, especially in real environments; however, when possible, it can provide information not identified using other assessment means (Datta, Bann, Mandalia, & Darzi, 2006). Procedures completed in a simulated environment can facilitate this functional assessment, for example, pressure testing a bowel anastomosis (Egle, Malladi, Gopinath, & Mittal, 2015) to examine the spacing, tightness and security of the sutures. This functional assessment is dependent on the model used, Egle and colleagues used porcine
intestine, whereas the latex tube used by Shah, Munz, Manson, Moorthy, and Darzi, (2006), would leak unrealistically and thus not provide valuable feedback.

Defining a task to facilitate deliberate practice requires deconstructing the activity into measurable and therefore assessable elements. Assessment tools can be used to both assess the level of performance and guide feedback on deficient areas. The utility of the feedback depends on the meaningfulness of the information provided. Assessment tools may focus on different aspects of the same performance and may be used to provide complementary information, such as the OSATS checklist and GRS. The widespread use of OSATS affords the opportunity to compare results across studies. It was for that reason that the OSATS was used as the assessment tool for my own studies.

**Highly motivated individuals: how personal motivation influences performance**

The second principle of deliberate practice is motivation of individuals to improve. Motivation originates from an individual’s psychological needs that he/she desires to have met (Deci & Ryan, 1985). These needs, or rather the discrepancy between the need and the individual’s current reality, generate the motivation required to change. Activities that can fulfil these needs become intrinsically important to the individual. From an educational perspective, educators have harnessed this motivation through self-directed learning activities to drive learning (Knowles, 1975). Thus, medical educators have two essential tasks: the primary task being to train physicians who are self-directed, life-long learners, and the second to provide the clinical foundation to support this independent learning, as Dr Dave Sackett (Smith, 2003) suggests that clinical knowledge moves so quickly that what was taught at medical school may no longer be relevant five years later. What is the theoretical rationale for promoting self-directed learning?
The use of self-directed learning in medical or other educational contexts assumes that the individual’s motivation is the key driver of his/her learning. The psychological needs according to Self-Determination Theory (SDT) (Deci & Ryan, 1985) are: autonomy (self-actualization), relatedness (belonging) and competence (esteem). These needs, or rather the discrepancy between the need and the individual’s current reality, generate the motivation required to change. Activities that can fulfil these needs, become intrinsically important to the individual. From an educational perspective, if educators can develop tasks where these needs can be fulfilled, then the learner will be much more motivated to complete the task.

Self-Determination Theory defines motivation as the force that moves an individual towards meeting his/her psychological needs. Originally this force had been defined broadly based on its orientation to the individual or self. Force generated from within the individual, e.g. personal interest, has been defined as intrinsic motivation, and the converse, an externally applied force e.g. parental pressure, has been defined as extrinsic motivation. In response to a more nuanced understanding of motivation, (Ryan & Deci, 2000) modified SDT by acknowledging the presence of an amotivated state, and sub-categorizing extrinsic motivation based on the perceived locus of control. The locus of control – where the source of the behavior comes from, can be considered a spectrum. This spectrum moves from a purely external locus e.g. to avoid punishment; to primarily external, with some internal locus, e.g. motivated by guilt or obligation; to predominately an internal locus with some external, e.g. studying to get into university; and finally, to an internal locus, e.g. exercising for the intrinsic pleasure of the activity. Ryan and Deci classified these as: external regulation, introjected regulation, identified regulation and integrated regulation respectively. Re-framing these motivations based the locus of control, Ratelle, Guay, Vallerand, Larose and Senécal, (2007) grouped the motives into three
overarching themes: “Autonomous regulation (i.e. acting out of choice), Controlled regulation
(i.e. acting for reward or avoid punishment), and Amotivation” (p. 734). The relevance of such a
grouping is that autonomously regulated learners are more successful than their peers.
Autonomous learners demonstrate “higher academic achievement (e.g., Areepattamannil and
Freeman 2008; Areepattamannil et al. 2010; Gottfried et al. 2007), higher intellectual
performance (e.g., Gottfried and Gottfried 1996, 2004), higher self-esteem (e.g., Deci and Ryan
1995), greater persistence (e.g., Vansteenkiste et al. 2004, 2006), [and] less academic anxiety
(e.g., Gottfried 1982, 1985, 1990)” (Areepattamannil, Freeman, & Klinger, 2011, p. 429). The
limitation of SDT is that it defines motivation as essentially a bidirectional vector: positive or
negative, and the overall direction of the motivational force is calculated by the sum of the
different motives. Task selection is based on the task with the highest motivational force. In
reality, one less appealing task might be selected over another, something that SDT is unable to
account for.

In undergraduate medical education, students with higher levels of autonomous
regulation have been found to have higher academic performance (Williams & Deci, 1996). The
cultural environment of the learning institution can impact the degree of students autonomous
orientation (Williams & Deci, 1996), with the autonomously supported faculty encouraging
students to become more autonomous learners. In residency training, the amount of autonomous
learning demonstrated by residents appears to impact amount of independence given by faculty,
with passive residents being given less autonomy than those residents showing higher levels of
motivation (Biondi et al., 2015). This highlights the impact faculty or trainers have on learners
motivation and as such recruitment and retention of such staff should be a priority (Lyness,
Lurie, Ward, Mooney, & Lambert, 2013). Although little has been identified in the literature
regarding physician learners, patient weight-loss and smoking cessation rates were higher, when physicians used an autonomy supportive approach to patient healthcare (Patrick & Williams, 2009).

Expectancy-Value Theory (EVT) unlike SDT, focusses on the task itself. EVT explores the rationale for why learners choose certain tasks, why they persist during hardship, and how this rationale affects overall performance (Eccles & Wigfield, 2002). Expectancy can be defined as the beliefs about how well an individual will perform during a future task. Expectancy is influenced by both personal and societal perceptions. Personal perceptions include: difficulty of the task, previous success, and competence to complete task. Societal perceptions include: parental or educator expectations, and societal value of the task.

Assigning a personal value to the task is dependent on four factors. Attainment – “the personal importance of doing well” (Eccles & Wigfield, 2002, p. 119) relative to the individual’s sense of identity. For example, getting an A in English because I view myself as a “smart” person. Intrinsic value – the pleasure obtained from tasks which the individual finds interesting for his/her own sake. Utility cost – how useful a task is to an overarching goal, for example, learning guitar to become a famous musician. Finally, cost – the negative elements of performing a specific task, for example: the time requirements, performance anxiety, or loss of opportunity to engage in other competing tasks. Since EVT’s inception in 1983 it has been empirically validated via numerous studies (see Eccles & Wigfield, 2002, for review).

These two main motivational theories highlight that by encouraging learners to develop a self-ascribed, or autonomously based motivation they are more likely to achieve the desired goal. However, motivation alone is insufficient to achieve success, it must also be incorporated into a cognitive cycle of planning, performing and reflecting on a task, that a learner conducts whilst
achieving his/her goal (Zimmerman, 2002). Described by Zimmerman (2002) in his self-regulation theory, this active self-assessment process is heavily reliant on the individual’s self-assessment skills, that determine whether the learner continues to strive toward, or modify or fail to achieve his/her goal. Those students who use self-regulated learning approaches have been found to have higher academic performance (Nicol, 2009; Tagawa, 2008), deeper understanding of topics (Räisänen et al., 2016), higher academic efficiency in Doctoral students (Kelley & Salisbury-Glennon, 2016) and greater life-long learning (van de Wiel & Van den Bossche, 2013; van de Wiel et al., 2011). There is clear evidence then from both conceptual and empirical perspectives that high motivation and associated self-regulation skills enable learners to achieve the highest levels of performance, which is in keeping with Ericsson’s principles of deliberate practice.

However, when examining skills outside of medical trainees traditional scholastic training, it has been identified that self-assessment, and the other aspects of Zimmerman’s self-reflective phase are often poorly developed (Davis et al., 2006), especially in those students who struggle academically. Building a curriculum to teach a technical skill using a deliberate practice framework, educators must therefore bolster the autonomously regulated motives of the learner, whilst also teaching and reinforcing those self-regulation skills necessary for students to achieve their learning goals.

**Repetition: facilitating repetition in medical education through simulation-based training**

The third principle of deliberate practice is the opportunity for repetition. Repetition allows the learner to become familiar with the task itself, and the necessary skills to perform it. In DP, this repetition is deliberately focused on areas of learner weakness and areas of complexity, rather than practicing the whole task, which can lead to replication of the same
errors. Traditionally, the exclusive pathway to teaching technical skills was on real patients during their required operations. Opportunities to practice a skill were thus dependent on patient volume and case-mix. This patient volume and case-mix may have been adequate for frequent low-stakes tasks but may have been insufficient in preparing trainees for rare high-stakes situations (Ziv et al., 2006). This approach to practice also required learners to practice the entire task, rather than focus on specific areas of concern. These concerns along with the other impacts on surgical training described in the introduction have all resulted in educators seeking alternative opportunities for learners to repeat the task or skill. Simulation based training has arisen to be the primary means of providing opportunities for repetition.

Simulation requires appropriate task alignment, in that simulations must able to adequately replicate the task with appropriate fidelity to meet the desired learning outcome. For example, using a low-fidelity model such as a Resusci-Anne (“Resusci Anne QCPR,” 2018) to teach the mechanics of CPR. A mismatch in either will result in either an inability to address the learning outcomes, e.g. if advanced life-support skills were being taught, or poor resource utilization, with an expensive model being used when a cheaper alternative is available.

For simulation to be used preferentially over traditional apprenticeship-based learning modalities, it must also fulfil two major requirements: simulation implementation must be based on appropriate learning theory, and simulation as a teaching tool must provide a clear benefit over existing methods. The learning theories most commonly applied to simulation are the constructivist view of learning by Dewey and Bruner (Pasquale, p. 52 from Levine et al., 2013). The constructivist view of learning suggests that new learning is built/constructed upon the foundation of previous learning using linkages or associations. Other theorists such as Kolb and Schön (as cited in Mann, Gordon, & MacLeod, 2009) have then sought to identify how these
constructs occur, through their experiential learning and reflective practice theories respectively. These theories infer that learning is not passive or automatic, learning is what occurs in response to the stimuli. The stronger the engagement the greater the learning, that “experience imprints knowledge more readily than didactic or online presentations alone” (Pasquale, p. 52 from Levine et al., 2013).

“Simulation holds the potential to operationalize the constructivist framework, in that it provides active engagement with the content, coupled with application to real-world activities” (Pasquale, p. 52 from Levine et al., 2013). Well-developed experiential learning activities such as simulation build, as described by Kolb, upon an event. Learners engage with the event, actively reflect upon it and then experiment as they apply this knowledge to the event, or an alternative. Only through this active reflection and experimentation does deep learning occur. Simulation seeks to place a part of the real-world within a learning environment (Jones et al., 2015). This containment then provides learners with the opportunity of time and space to reflect and experiment that may not be accessible in a real-life environment. For the educator, this containment also focuses the educational purpose of the activity. Specific learning objectives can be developed prior to the activity, and educational tasks developed to achieve them (Burke & Brodkey, 2006). Reflection on the activity as highlighted by Schon, is fundamental for the translation of the event experience into learning. Despite reflection being such a crucial component of experiential learning, evidence from even experienced learners demonstrate that this metacognitive skill is often poorly developed (MacKenzie, 2014; Ryan, 2010).

Aside from having a theoretical justification for its application, experiential learning activities through medical simulation must also provide clear advantages over experiential activities found in the traditional apprenticeship model. Ziv, Wolpe, Small, and Glick (2006)
highlight the four domains where there are clear advantages to using simulation over patient-based learning: patient care, standard of training, skills-based evaluation, and error management and patient safety. One of the ethical principles underlying patient care is to do no harm. The tension arises between providing training opportunities for trainees, that may place patients at risk of harm with no direct benefit for the patient. Simulation removes much of the risk of these unintentional harms from patients. Using an intermediate simulation-based learning stage before direct patient contact ensures that trainees are at a higher level of skill, thereby reducing risk of unintentional harm to patients due to inexperience. Simulation improves the standard of training by providing reproducible structured technical and non-technical training, that is not dependent on available patients within the health facility. Simulation allows uncommon but critical clinical situations to be practiced in a planned and structured way, something the traditional approach is unable to do, as it is entirely dependent on available patients. The third advantage is that simulation can be used to assess both non-technical and technical skills of the trainee in a detailed manner, without impairing the therapeutic relationship, for example end of life discussions. The final advantage, patient safety also utilizes the principle of do no harm. Mistakes are a powerful learning opportunity, simulation allows mistakes to progress to their natural conclusion without any harm occurring to the patient. In the traditional model, errors were required to be corrected early to prevent such harm occurring. Additionally, trainees can repeat the situation and experiment with their own technique, which is consistent with experiential learning theory. This lack of consequence also promotes positive non-blame reflection of the error. In a real environment, this open discussion is often negatively influenced by concerns of litigation and other punitive consequences to the learner or team.
Although these advantages provide clear justification for the implementation of simulation in healthcare training, there are also a number of significant costs which limit its use. Current evidence suggests that feedback on simulation is best provided by experienced practicing clinicians (Porte et al., 2007). Experts providing this feedback, particularly complex surgical tasks are expensive, either as a direct cost to the training provider or in lost income to the clinician. Post graduate training occurs in a working health facility, in which trainees and supervisors are employed to provide patient care. Health facilities must then bear the opportunity cost of reduced service delivery that occurs when staff are doing internal training. Health facilities linked to educational providers such as universities, may have defined remuneration package to offset this cost, but for other facilities, this cost must be borne directly by the facility. Simulation is also an expensive method of delivering education to a large group of learners (Levine et al., 2013). Higher educator to trainee ratios are required to enable the appropriate level of personalized feedback to be provided. In addition, simulation equipment is expensive. Aside from the direct costs of the equipment itself, are the ongoing costs of facilities, support staff and disposables required to ensure that the equipment can be used for education. The final limitation of simulation is the adequacy of the model and associated assessment methods used by the educators. For example, a virtual reality task trainer will not be able to provide the same haptic feedback that a real operation can.

Recognizing the limitations of simulation allows educators to be more selective over when simulation is utilized and how it can be delivered in the most cost-effective way possible. This understanding allowed (Zevin et al., 2018) to develop a structured simulation-based curriculum for teaching laparoscopic bariatric surgery. The major elements they identified were as follows: 1) an initial knowledge-based learning period to prepare trainees for the procedure; 2)
deconstruction of the procedure into relevant tasks with concordant development and validation of assessment tools; 3) laboratory training to proficiency of those pertinent tasks on a suitable model; 4) transfer of those skills to the real environment with further confirmation of trainee proficiency; and 5) validation of the achieved skills with appropriate permissions, as well as validation of the teaching methodology.

Simulation based training creates the opportunity in which learners can repetitively practice specific elements of a task, in a safe and consequence-free environment. This educational focus, when combined with the other aspects of Ericsson’s DP model, allows the learner to rapidly improve their performance beyond what is possible by mere repetition alone. However, simulation is not the universal panacea for surgical education. Simulations significant limitations mean that without strategic implementation into the broader surgical curriculum, particularly in the early stages of skill acquisition, educator and learner alike may become indifferent to its use.

**Feedback: using feedback to improve performance and assist learners in building their self-regulated learning skills**

The final and most crucial principle of DP is that of the provision of feedback. Ericsson’s research highlighted the importance of feedback (Ericsson, 2008) most notably when given by an expert familiar with the goals and aims of the task. This section examines the rationale for feedback, the skills required to provide feedback, and the effectiveness of alternative sources of feedback such as peer or self-feedback. The section will then conclude with how feedback from any source may be integrated into a physicians clinical practice using reflective practice.

Hattie and Timperley (2007) describe in their model of effective feedback (see Figure 1) that educators must have clear goals and objectives to which learners can aspire. This fits well
with Ericsson’s first principle of DP, that of a well-defined task. A well-defined task provides clear guidance on the sub-components of the task, that can be categorized into learning objectives and sub-goals. These goals should be complex enough to provide challenge, but also focused and manageable. Advantages of using learning objectives and other sub-goals are that they break down the task into manageable components, act as a guide for delivering feedback and can also be used as a method of assessing the overall performance of the task.

![Diagram](image)


In addition to the understanding of the content required to deliver feedback, it is necessary to recognize that feedback does not occur outside of the educational context in which it was requested. It is the result of this interplay between the environment and the content of the feedback, that feedback is either accepted, rejected or modified by the learner (Hattie & Timperley, 2007). Despite the risk of this feedback being incorrectly applied, interpreted or
ignored, Hattie and Timperley (2007) identified that it was one of the top ten influences on achievement.

Some forms of feedback are more effective than others, particularly those that “provide cues or reinforcement to learners; are in the form of video-, audio-, or computer-assisted instructional feedback; and/or relate to goals” (Hattie & Timperley, 2007, p. 84). Rewards, punishments or praise are the least effective methods of providing feedback, due to their lack of task relevant information (Zhang & Bednall, 2016). This is consistent with motivational theory (Deci & Ryan, 2000; Eccles & Wigfield, 2002) where extrinsic motivation has been found to negatively affect performance (Areepattamannil et al., 2011; Heikkilä & Lonka, 2006; Vallerand & Blissonnette, 1992).

Using video as part of the feedback process is, however, not without limitation due to either logistical or implementation problems. Logistical limitations are associated with the cost of acquiring cameras and the consequences of any failure in the recording process. It may be possible to mitigate the cost of purchasing cameras by borrowing or hiring cameras. Incomplete video recording can occur and are a consequence of either inadequate battery life or storage capacity. Another logistical limitation identified is associated with the sharing of the footage. Large file sizes can be difficult to transfer between devices, which may impact the ability of others to review the footage.

Implementation problems revolve around the time required to review the video. The time taken to review unedited videos may be significant depending on the length of the procedure, for example it took blinded experts between 30 to 40 minutes to review a 120-minute video (Zevin et al., 2013). Increasing the speed of review (Dath et al., 2004) has been shown to be effective, but increases the apparent smoothness and speed the procedure, which may affect scoring
(Nickel et al., 2015). Using edited videos to provide focus on areas of concern is beneficial (Datta et al., 2006), however still incurs a time cost, and potentially misses errors. Video review also delays the identification and correction of any issues that occur, as these issues are only detected after the procedure has been completed.

A lack of access to video recording devices does not preclude peer-feedback from being used. Indeed, having a peer available to observe the procedure live does allow contemporaneous feedback to occur, so that issues are immediately addressed before learners move on to the next component of the procedure. This method does require a peer to act as an observer for the whole procedure, preventing them from practicing themselves and therefore increasing the training time required. However, based on Kolb’s learning cycle (Levine et al., 2013), learners benefit from the opportunity to observe and reflect on an experience. Without the availability of video-review, faculty or coaches are left to attempt to replicate the identified issue to help the learner observe and reflect on the incorrect technique. If video were available, this would also promote Kolb’s forth stage of learning, that of experimentation, as video would provide learners the ability to compare the results of their experimentation with earlier performances.

Feedback is information that is provided to the learner to improve the learners’ performance or understanding toward a defined ideal. This therefore requires an intimate knowledge of the ideal or standard to which the learner must meet. This intimate knowledge can be provided by an expert or coach, or after ‘years of daily practice, the aspiring expert performer’ can “become able to monitor their performance so they can start taking over the evaluative activity of the teacher and coach” (Ericsson, 2008, p. 991). Feedback is then directed in supporting positive or stopping negative actions that improve performance or understanding.
Feedback should also aim to foster learning on multiple levels (Burke & Brodkey, 2006), providing students with the opportunity to develop the higher-level cognitive skills necessary to evaluate their own or others work. This evaluative judgement: “the capability to make decisions about the quality of work of self and others” (Tai, Ajjawi, Boud, Dawson, & Panadero, 2018, p. 471) is the foundational component of providing feedback. If learners can acquire or develop this skill of evaluative judgement, then they are able to provide relevant feedback. Further training on feedback delivery can build upon the learner’s understanding of quality.

A common method used to engage students in the evaluation process, thereby building students’ evaluative judgement, is through self-assessment and peer feedback. In peer feedback, neither student claims expertise, thus there is an absence of the power dynamics found in traditional expert based feedback models. The feedback between equal peers can be described as a learning dialogue: an active two-way dialogue form of feedback that enables free sharing of ideas and concerns (Askew, 2000). A learning dialogue allows “feedback and reflection [to] become entwined, enabling the learner to review their learning in its context and relate [it] to previous experiences and understandings” (Askew, 2000, p. 13). This sharing of ideas and concerns reduces the barriers to engaging in the evaluation process. As an example, peer feedback in a technical design project in an Engineering context was found to both improve the technical quality of the product as well as provide students with insights into their own evaluative judgement and learning (Nicol, Thomson, & Breslin, 2014).

Peer assessment “enables students to better self-assess themselves as some skills are common to both peer and self-assessment” (Liu & Carless, 2006, p. 281). These self-regulated learning styles are associated with higher academic performance (Nicol, 2009; Tagawa, 2008), deeper understanding of topics (Räisänen et al., 2016), higher academic efficiency in Doctoral
students (Kelley & Salisbury-Glennon, 2016) and greater life-long learning (van de Wiel & Van den Bossche, 2013; van de Wiel et al., 2011).

When looking more broadly at the literature on processes that require learners to evaluate some aspect of themselves or others, the following terms are used to describe very similar activities: self-assessment, self-reflection, review or reflective practice. What these terms are describing is not the evaluative process or the activities themselves, but rather the purposes for which these evaluations are occurring. For example, Desjarlais and Smith (2011) describe self-reflection as being an internally driven activity used to “gain insights into themselves” (p. 10) whereas self-assessment is used to “improve future performance” (p. 10). Both self-assessment and self-reflection involve an iterative process of: identification, assessment against a standard, interpretation of the assessment, and application through planning and implementation; yet this commonality between the two is overlooked within the educational literature. Hence this thesis uses the term evaluative judgement to examine the process of review rather than the equivalent but potentially misleading self-assessment, self-review or self-reflection.

Having now discussed the benefits of feedback, I will turn now to the influence of feedback has on a professional career, through the concept of reflective practice. I will define reflective practice and will conclude with a discussion about the benefits and implementation of reflective practice in a medical curriculum.

Mann, Gordon, and MacLeod's (2009) description of reflective practice adds an additional element to the task assessment and the personal introspective reflection debate: time. The authors suggest that when these evaluative activities are performed regularly an “understanding of one’s personal beliefs, attitudes and values” (p. 596) and “critical reflection on experience and practice that would enable identification of learning needs” (p. 596) to ultimately
allow the learner to develop into “a professional who is self-aware, and therefore able to engage in self-monitoring and self-regulation” (p. 596).

If reflective practice is so beneficial, why has it not been fully inculcated into medical education processes? Three factors commonly identified in the reflective practice literature have been found to inhibit the implementation of these practices: the time cost, method of implementation and faculty training. Medical education is full of competing curricular and clinical activities, when examining existing student and resident reflective practices, studies (Larsen, London, & Emke, 2016; Peshkepija et al., 2016) identified that both groups found it difficult to find the time required to engage in reflection. Even when reflective activities are introduced, the method of implementation appears to have a significant impact. MacKenzie (2014) and Peshkepija et al. (2016) highlighted that learners felt these reflective practice tasks: e-portfolios and its associated reflections, and reflections on surgical performance, were imposed upon them by the training body, and as such could see little value in their professional careers. However, as Peshkepija et al. (2016) went on to comment, when provided with explanation and training, the quality of the responses as well as the perceived benefits increased. The final factor inhibiting implementation is when faculty themselves have no experience of, are not adequately trained in how to perform reflective practice or themselves feel these reflective practice tasks are irrelevant (Anderson, 2012; Larsen et al., 2016). This role-modelling significantly impacts the effectiveness of any educational intervention, but even more so for non-traditional tasks like reflective practice (Chelliah & Arumugam, 2012). Although these limitations are significant the literature is replete with suggestions on how to effectively implement reflective practice.

For reflective practice to be valued by the medical community, changes at all levels are required. The primary drivers of this change are the professional medical Colleges. Without the
Colleges advocating and mandating for these reflective practice activities, large scale changes are unlikely to occur. Some Colleges have begun this change process, for example, making reflection a mandatory part of continuous medical education (Row et al., 2018; Royal Australian College of Surgeons, 2017), or requiring this as part of medical licensure (Heeneman & Driessen, 2017). At a lower level, training programs have two roles to play in changing the value of reflective practice. Firstly, they need to consider implementing reflective practice activities within their training curriculum (Ryan, 2010; Wald, Borkan, Taylor, Anthony, & Reis, 2012). Secondly, programs need to educate faculty and learners of the benefits of reflective practice activities (Peshkepija et al., 2016), as lack of learner education was found to be a significant inhibitor to student engagement with electronic portfolios (MacKenzie, 2014). The curriculum developed for the training programs must be developed by educational experts who have a clear understanding of the methods by which reflective practice activities can be appropriately implemented within the training curricula to maximise learner engagement (Edwards & Thomas, 2010; Russell, 2005). At the lowest levels, faculty need to model reflective practice behaviours with learners to provide learners with clear examples of how such behaviours can be incorporated into clinical life (Chelliah & Arumugam, 2012). Motivated learners need to also contribute and agitate for reflective activities to be built into their training curriculum.

To effectively implement reflective activities three essential concepts are necessary: Integration, facilitation and training in reflective practice, and time to engage in reflective practice. The first concept is ultimately the key on which all others can hang; integrating reflective activities into the educational activity, rather than having them seen as a stand-alone tasks (Russell, 2005). As Edwards and Thomas (2010) argue “reflection and practice are indivisible aspects of the same process” (p. 406), so to artificially separate reflective practice
“The problem with this artificial separation is that the focus on teacher education shifts away from supporting new teachers in solving their pedagogical problems towards monitoring their acquisition of a rubric of illusory reflective attributes” (p. 410). If tasks naturally include reflective components, that are relevant to task, then learners will engage naturally in reflection. Russell (2005) suggests that if the reflection is well integrated, you need only remind the learner that they have already engaged in reflection. This positively reinforces the value of reflection, rather than risking negative attitudes to mandatory reflections once the task is finished and learners have disengaged.

The second requirement for implementing reflective practice is allowing time to engage in reflective activities. Asking questions is helpful (McNicol, Lewin, Keune, & Toikkanen, 2014), but learners need time to achieve deep or critical thinking. This can be difficult to achieve, but as highlighted above, when reflective tasks are integrated into an educational activity, space is created within the curriculum for reflection to occur (Larsen et al., 2016).

The final concept of implementation is that like any other educational concept, reflective practice is a skill that can be taught (Wald, 2012; Wald et al., 2012). Therefore, by teaching the learner the rationale of reflective practice, the skills that are involved, how to apply these skills to a medical context using good role-modelling by facilitators, then effective reflective practice will take place (Peshkepija et al., 2016; Ryan, 2010). Although this seems rather counter-intuitive to an integrated reflection, much of this learner training can be accomplished through the appropriate curriculum design or facilitator role-modelling. For example, dedicated instruction time on how to provide feedback, or facilitators articulating the reflection process by probing learners about what they are doing within the task, how this may affect themselves or others, or
Feedback is an essential component not only of deliberate practice, but any educational activity that desires that learners build new understanding or skills. For feedback to be effective, the provider of feedback requires training in both the task and method of delivery. Although faculty already have the technical expertise of the task, it is possible for the learners to gain this evaluative judgement to facilitate peer and self-feedback. Reflective practice, a life-long attitude and an expected approach to professional learning consist of skills and behaviors that require practice. Early exposure to these evaluative skills such as reviewing performance, assessment and feedback, would engender novice learners to these reflective practices, through deliberate training, facilitation and integration into the learning activity.
Chapter 3

Technical Skills Study

Abstract

Purpose
Video feedback and faculty feedback has been shown to improve surgical performance; however, consistent access to faculty is challenging. We studied the utility of structured peer-feedback (PF) compared to faculty-feedback (FF) during acquisition of basic and intermediate surgical skills.

Methodology
Two randomized non-inferiority trials were conducted with first (n=30) and second year (n=29) medical students learning skin-lesion excision and closure (S), and single-layer hand-sewn bowel anastomosis (B), respectively. Five attempts were performed. PF participants used an Objective Structured Assessment of Technical Skills tool to guide feedback. Blinded raters assessed video-recorded performance, time and Integrity of the completed task were also assessed.

Results
For both tasks performance by PF was comparable to FF (P=0.111). Both groups improved significantly: performance (B:P<0.0001, S:P=0.035), time (B:P=0.043, S:P<0.0001) and integrity (B:P<0.0001, S:P<0.032).

Conclusion
Structured peer-feedback is equivalent to faculty-feedback in the acquisition of basic and intermediate surgical skills, giving students freedom to practice independently.
**Introduction**

Postgraduate surgical education has been required to adapt to several significant societal shifts in recent years. The introduction of restricted working hours, patient safety concerns, economic and efficiency drives from hospitals have all affected the accessibility of residents learning technical skills. Alternative methods of training have subsequently been implemented to augment the traditional learning that occurs in the operating room (McGaghie, Issenberg, Petrusa, & Scalese, 2006). In particular, during the early stages of surgical training, surgical boot camps and simulation laboratories have been shown to be effective alternative approaches to teach fundamental surgical skills (Fernandez et al., 2012; Levine et al., 2013).

Irrespective of the strategies employed, it is the adherence to the evidence-based principals of educational practice that enabled these initiatives to succeed. One of these principles is the provision of timely and high-quality feedback during skill training. Anders Ericsson, through his research on deliberate practice (Ericsson, 2008), has been able to demonstrate that provision of high quality feedback during practice is the only way to achieve expert level of performance. Ericsson’s research highlighted the importance of feedback (Ericsson, 2008) provided by an expert; however, in the medical context, consistent access to faculty for feedback remains a challenge (Zevin et al., 2018).

To overcome some of the challenges with faculty participation in providing feedback, alternatives such as self-assessment strategies or peer-feedback have been studied, with results generally favoring expert feedback over either self or peer approaches (Jamshidi, LaMasters, Eisenberg, Duh, & Curet, 2009; Nesbitt et al., 2015; Porte et al., 2007).

Although these studies have generally compared feedback from trained experts with untrained novice individuals and peers. Studies that have provided more guidance and training to
the peer-based feedback groups have in fact demonstrated positive outcomes from peer-based feedback (Hoogenes et al., 2015; Vaughn et al., 2016). This additional guidance and training provided students with the structured framework necessary to critically review their own, or their peer’s performance, thereby gaining insight into the task and guiding improvement in their performance.

The purpose of our study was to build on prior research by further developing the concept of structured guidance. Structured guidance is defined as the reflective practice interventions that are provided to students in order for them to knowledgeably examine their performance of a task. For example, training students to perform an assessment of the procedure, or to complete a self-reflection. These reflective tasks that are used as part of structured guidance have been shown to improve diagnostic accuracy (Mann, 2016). To achieve this outcome of improved diagnostic accuracy, these reflective tasks need to be appropriately introduced into the training curriculum. When poorly implemented, these tasks have been resented by students (Anderson, 2012; Larrivee, 2008), for example, students that are requested to reflect on their performance, but are not provided with the requisite skills or understand the expectations of the reflection activity (Russell, 2005). Our primary research objective was to investigate whether students’ performance improves with structured peer-feedback as compared to traditional faculty feedback for both basic and intermediate level surgical tasks. Our secondary research objective was to explore if the inclusion of these reflective activities during technical skill practice was perceived to be beneficial by study participants.

**Methods**

Two prospective randomized non-inferiority controlled trials were conducted with first and second year medical students at Queen’s University, Canada examining technical skill
acquisition under two types of feedback conditions. Structured peer-feedback was the intervention and faculty feedback the control. Study 1 examined acquisition of intermediate level technical skill - a single layer interrupted handsewn bowel anastomosis using cadaveric porcine small bowel. Study 2 examined acquisition of basic level technical skill - a skin lesion excision and suture closure using cadaveric porcine skin. The objective of each study was to investigate whether provision of structured peer-feedback resulted in non-inferior technical skills as compared to provision of faculty feedback. All participants in both studies were provided with relevant theory and shown a video demonstration of the task they were required to perform, prior to commencing their individual video recorded practices.

**Feedback types.**

Faculty feedback was based primarily on the three assessment domains used in this study: the technical elements articulated in the objective structured assessment of technical skills (OSATS) assessment tool (Martin et al., 1997), the time taken to complete the procedure, and the observed results of the functional assessment of each final product, as described below. The OSATS assessment tool contains a task specific checklist and global rating scale that examines different aspects of technical performance. Faculty members discussed future goals and areas for improvement during the feedback session. Faculty members were trained in how to use the OSATS tool for feedback prior to the commencement of the study. Feedback was individualized and based on direct observation, with a maximum faculty to participant ratio of 1:8.

Participants in the peer-feedback groups provided feedback to each other based on the reflective activities taught to the groups prior to commencing their individual practices. This was achieved by providing structured guidance in how to use the OSATS tool for both assessment and feedback purposes using frame of reference training by assessing a video of a poor
performance. Additionally, self-reflection prompts were added into the OSATS documentation to assist in the feedback process, for example, asking assessors to justify why they scored the way they did (Appendix A). After participants in the peer-feedback group were satisfied that they understood how to provide peer-feedback using these reflective tools, they commenced practicing the technical skill individually. A self-selected peer from peer-feedback group, reviewed and assessed his/her video using the OSATS tool immediately post completion of the task, reviewed the time taken to complete the procedure, and the observed results of the functional assessment of the final product. Based upon the OSATS score, the time taken and functional assessment results, the peer and study participant collectively developed goals for the participant’s subsequent practice attempt.

**Study 1: Bowel Anastomosis.**

**Participants, inclusion, and exclusion criteria.**

Participants were recruited from second year medical students at a Canadian University who were participating in a summer surgical skills training program. Participants were recruited as a convenience sample via email invitations. Participants were excluded from participating if they could not dedicate enough time to complete a total of five bowel anastomosis, and if they were unable to suture or perform an instrument tie for a surgical knot.

**Randomization.**

Closed envelope technique using a sample random number generator.

**Description of learned technical skill.**

Participants were asked to anastomose two segments of transected 25cm cadaveric porcine small bowel using a single layer interrupted hand-sewn end to end suture technique (Egle, Malladi, Gopinath, & Mittal, 2015; Martin et al., 1997; Shah et al., 2006) using 3.0 silk or
3.0 Vicryl®. Participants practiced the anastomosis task a minimum of five times over four days, receiving concurrent feedback throughout, with the method of feedback determined by their randomization. Each attempt was video recorded with cameras focused exclusively on the operative field, de-identifying study participants. At the completion of each anastomosis, a pressure test using a column of water was conducted until leakage occurred (Figure 2). This test provided additional visual feedback to the participants about the quality of their constructed anastomosis.

Figure 2 Anastomotic leak test

Study 2: Skin lesion excision.

Participants, inclusion, and exclusion criteria.

Participants were recruited from first year medical students at a Canadian University using the same methodology as Study 1. Participants were excluded if they were not able to commit to completing five attempts and if they were unable to suture or perform an instrument tie for a surgical knot.
**Randomization.**

Closed envelope technique using a sample random number generator.

**Description of learned technical skill.**

Participants were asked to excise a 3-8mm diameter skin lesion, painted onto a cadaveric piece of porcine skin, using 3:1 ratio elliptical excision with a minimum 5mm margin. The skin defect was closed using an interrupted single layer suture technique using 3.0 Prolene® or 3.0 Monocryl®. A functional assessment of the strength of skin closure was performed using a 5kg accumulated weight displacement test, see figure 3 (Westlake, 2012). This test provided participants additional visual feedback about the quality of their skin closure. Failure of the functional test was defined as a wound edge displacement of 5mm or if the suture knots failed or tore through. Participants practiced the task a minimum of five times over four days. All practice attempts were video recorded in a de-identified manner.

*Figure 3. Skin excision and closure stress test.*

**Baseline demographic data.**
Participant age, sex, current year of medical school training, participant expectations and perceptions of learning these skills, and previous technical expertise were obtained at the commencement of the study (Appendix B). Technical expertise was subdivided into suturing (how many suturing opportunities had they had previously, e.g. three previous suturing workshops), and exposure to the specific task. Exposure was defined as having seen the task or assisted/Performed the task.

**Assessment of technical performance post completion of training.**

Technical performance was assessed across three domains: OSATS assessment of all videos, time to complete the procedure (minutes), and the outcomes of the functional assessments of the completed product as described above. All videos from both control and intervention groups were collated, randomized, using a Microsoft Excel® random number generator, and provided to two trained and blinded reviewers (raters) for assessment using the OSATS tool. Part of rater training included reaching a consensus on interpretation of technical performance prior to assessing the videos of study participants. Raters were asked to submit their own videos of their own performance. One of the videos was then used for the video demonstration and rating. Raters were able to fast forward the video according to their preference (Dath et al., 2004). Videos with assessment scores that differed by at least 9 points between raters were re-assessed by both raters to achieve consensus, raters were blinded to their previous scores to prevent forced consensus.

Post study survey examined participants views on the study, the role of feedback, their preferred feedback method, as well as their experience using the reflective activities taught in the study (Appendix C). Quantitative questions used a Likert-type scale (1-strongly disagree, 3-
neutral and 5-strongly agree). Short answer question responses were collated and examined for emergent themes.

**Sample size calculation.**

A sample size of 13 participants in each group was required for each study, based on a non-inferiority power calculation (“Sealed Envelope | Power calculator for continuous outcome non-inferiority trial,” n.d.) using standard deviation of 3.0 (Olson, Becker, McDonald, & Gould, 2012), α of 0.05, power of 80%, and δ of 3.0. Both studies were conducted in accordance with university ethics requirements (Queen’s University IRB no: 6020826) and were conducted at School of Medicine Simulation Centre.

**Statistical analysis.**

Demographic data was analyzed using descriptive statistics. OSATS scores and final product analysis scores are reported as means (standard deviations). Between group comparisons of continuous data was performed using a mixed ANOVA. Within group comparisons were performed using a mixed ANOVA with Bonferroni correction for repeated measures. Inter-rater reliability was calculated using an intraclass correlation two-way random effects model. Survey response comparisons were assessed using a Mann-Whitney U test. Data was analyzed using SPSS® version 24.

**Results**

**Study 1: Bowel anastomosis.**

Twenty-nine second year medical students (12 males, 17 females, age 25.4(3.5) yrs.) participated in this study. Participants reported previous technical experience of 9.0(5.6) exposures to suturing. Six participants had previously seen a bowel anastomosis and one participant previously assisted in the creation of more than three anastomoses. There were no
differences between groups in baseline experience ($p=0.291$). Baseline and fifth attempt assessment scores are shown in Table 1.

Table 1. Bowel anastomosis: performance at baseline and fifth attempt.

<table>
<thead>
<tr>
<th></th>
<th>Number of participants</th>
<th>Number of attempts</th>
<th>OSATS</th>
<th>Time (mins)</th>
<th>Pressure (cmH20)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Baseline</td>
<td>5th attempt</td>
<td>Baseline</td>
</tr>
<tr>
<td>Faculty</td>
<td>15</td>
<td>5.8 (1.1)</td>
<td>32.2 (6.1)</td>
<td>43.4* (3.2)</td>
<td>39.5 (10.8)</td>
</tr>
<tr>
<td>Peer</td>
<td>14</td>
<td>5.2 (0.6)</td>
<td>34.0 (5.2)</td>
<td>43.7* (3.6)</td>
<td>41.2 (12.6)</td>
</tr>
</tbody>
</table>

Note: *= significant group change from Baseline ($p<0.05$)

**Technical skill (OSATS scores).**

There was no statistically significant difference in technical skills at the completion of training between peer-feedback and faculty-feedback groups ($p=0.11$). Participants in peer-feedback and faculty-feedback groups demonstrated a significant within group improvement in their technical skills across all five attempts ($p<0.0001$; Figure 4). Similar results were also obtained for checklist and global rating components of the OSATS tool (Table 2). There was excellent inter-rater agreement for OSATS scores (ICC = 0.84; $p<0.0001$).
Figure 4. Bowel anastomosis: expert OSATS scores by attempt.

Table 2. Bowel anastomosis: OSATS scores subdivided into checklist and global rating scale.

<table>
<thead>
<tr>
<th></th>
<th>OSATS 1&lt;sup&gt;st&lt;/sup&gt; Attempt</th>
<th>OSATS 5&lt;sup&gt;th&lt;/sup&gt; Attempt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Checklist</td>
<td>Global Rating Scale</td>
</tr>
<tr>
<td>Faculty</td>
<td>15.1 (2.6)</td>
<td>17.1 (4.0)</td>
</tr>
<tr>
<td>Peer</td>
<td>16.0 (2.3)</td>
<td>18.0 (3.1)</td>
</tr>
<tr>
<td>Between groups</td>
<td>p=0.563</td>
<td>p=0.371</td>
</tr>
</tbody>
</table>

**Time to complete the task.**

There was no statistically significant difference between the peer-feedback and faculty-feedback group in the time required to complete the anastomosis (p=0.37). The time to complete an anastomosis improved over the five attempts for both groups (p=0.04; Figure 5). Although
statistically significant overall, pairwise differences between individual attempts were not significant.

**Figure 5. Bowel anastomosis: completion time by attempt.**

**Quality of final product.**

The quality of the anastomosis, as measured by the water pressure leak test, for peer-feedback and faculty-feedback groups was not significantly different at baseline or at completion of training ($p=0.47$). The quality of the anastomosis significantly improved over the five attempts for each group ($p<0.0001$; Figure 6).
Study 2: Skin excision.

Thirty first year medical students (12 males, 18 females, age 24.8(2.4) yrs.) participated in the skin excision study. Participants reported an average of 5.0(4.2) previous opportunities to practices suturing. Seven participants reporting observing more than three excisions or assisting in fewer than three skin excisions. There was no difference in baseline level of experience between the groups ($p=0.29$). Baseline and post-completion of training assessment scores for peer-feedback and faculty-feedback groups are reported in Table 3.
Table 3. Skin excision: performance at baseline and fifth attempt.

<table>
<thead>
<tr>
<th></th>
<th>Number of participants</th>
<th>Number of attempts</th>
<th>OSATS</th>
<th>Time (mins)</th>
<th>Force Pass Rate (5kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Baseline</td>
<td>5&lt;sup&gt;th&lt;/sup&gt; attempt</td>
<td>Baseline</td>
</tr>
<tr>
<td>Faculty</td>
<td>15</td>
<td>6.1 (0.9)</td>
<td>36.5 (7.2)</td>
<td>41.9* (5.3)</td>
<td>53.6 (13.8)</td>
</tr>
<tr>
<td>Peer</td>
<td>15</td>
<td>5.5 (0.6)</td>
<td>36.0 (7.1)</td>
<td>39.2* (8.7)</td>
<td>58.6 (16.3)</td>
</tr>
</tbody>
</table>

*Technical skill (OSATS scores).*

There was no difference in technical skill at the completion of training between peer-feedback and faculty-feedback groups ($p=0.56$). Participants in peer-feedback and faculty-feedback groups demonstrated a significant improved in their technical skills across all five attempts ($p=0.035$; Figure 7). Similar results were also obtained for checklist and global rating components of the OSATS tool (Table 4). There was good inter-rater agreement for OSATS scores (ICC = 0.72; $p=0.001$).
Figure 7. Skin excision: expert OSATS scores by attempt.

Table 4. Skin excision: OSATS scores subdivided into checklist and global rating scale.

<table>
<thead>
<tr>
<th></th>
<th>OSATS 1st Attempt</th>
<th>OSATS 5th Attempt</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Checklist</td>
<td>Global Rating Scale</td>
</tr>
<tr>
<td>Faculty</td>
<td>17.5 (3.3)</td>
<td>19.0 (4.6)</td>
</tr>
<tr>
<td>Peer</td>
<td>17.6 (3.8)</td>
<td>18.4 (4.3)</td>
</tr>
<tr>
<td>Between groups</td>
<td>P=0.940</td>
<td>P=0.719</td>
</tr>
</tbody>
</table>

*Time to complete the task.*

There was no statistically significant difference in duration of time required to complete the skin excision and closure between the peer-feedback and faculty-feedback groups at baseline.
or at the completion of training (P=0.49). There was a significant decrease in time required to complete the skin excision and closure over the five attempts for both feedback groups (P<0.0001; Figure 8).

![Figure 8. Skin excision: completion time by attempt.](image)

**Quality of final product.**

Participants in the peer-feedback and faculty-feedback groups achieved a 100% pass rate on the skin stress test by their third attempt - a significant improvement from the baseline score (P=0.032; Figure 9). The quality of final product was not significantly different between the peer-feedback and faculty-feedback groups (P=0.89).
Exploration of reflective activities (survey results).

In addition to the Likert-type questions shown in Table 5, participants were asked to describe what they learned from their involvement in the study. These short answer responses were collated and examined for common themes. Three main themes emerged. The first theme was the improvement in technical abilities. Participants (78.8%) reported that this study “improved (my) suture technique” and that “I learned how to tie better sutures and how far apart the sutures could be placed while still being effective”. The second theme was the value of practice. Participants (21.2%) reported that they could “confidently just go at a practice without the stress of trying to make it perfect from the start” and that “I got to experiment a lot with different techniques (bite sizes, how to grasp tissue)”. The final theme was feedback. Participants
(27.3%) reported that they “learned how to accept feedback and how to effectively reflect on mistakes as a method to improve learning” and “implementing feedback into manual techniques of suturing”. Of those responses, 75% focused on the technical aspects of giving or receiving feedback for performance enhancement, only 25% - all in the faculty group, commented on the broader metacognitive aspects of feedback and reflection. There was no significant difference in frequency of responses between the faculty or peer-feedback groups ($p=0.22$).

Table 5. Responses for Likert survey questions.

<table>
<thead>
<tr>
<th>Pre-study survey</th>
<th>Skin Bowel anastomosis</th>
<th>Skin excision and suture closure</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you rate your current ability to do the skill? (0-10)</td>
<td>1.4 (1.7)</td>
<td>$P^1=0.134$</td>
</tr>
<tr>
<td>How would you rate your confidence to learn to the skill? (0-10)</td>
<td>5.8 (3.0)</td>
<td>$P^1=0.434$</td>
</tr>
<tr>
<td>Post-study survey</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you think this study was successful in teaching you the skill? (1-5)</td>
<td>4.5 (0.5)</td>
<td>$P^1=0.601$</td>
</tr>
<tr>
<td>How comfortable were you learning the skill without faculty to guide you? (1-5)</td>
<td>3.6 (0.8)</td>
<td>N/A</td>
</tr>
<tr>
<td>Would you have preferred some faculty feedback in addition to the peer feedback? (1-5)</td>
<td>3.4 (1.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>How comfortable were you assessing your work with a peer? (1-5)</td>
<td>3.4 (0.8)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

1. Mann-Whitney U test between faculty and peer-feedback groups

**Discussion**

The results of our study demonstrate that peer-feedback was equivalent to faculty-feedback in acquisition of basic and intermediate technical skills (skin excision and bowel anastomosis) as assessed by OSATS scores of videotaped reviews of performance, time to completion of the task and quality of the final product. Furthermore, performance of participants
in both the peer-feedback and faculty-feedback groups improved significantly from baseline over the five attempts.

Our study used a process of deliberate practice (DP) for acquisition of technical skills. Deliberate practice, as identified by Ericsson, is based on the concept that it is the application of relevant feedback, combined with opportunities for repetition, that is necessary to achieve expertise (Ericsson, 2008). In order to provide relevant feedback, it is necessary to first adequately define the task itself. We purposely chose three assessment criteria to define each task. These criteria were chosen because they could be easily understood by the participants, and could be used to set individual learning goals (Porte et al., 2007).

We chose to use the OSATS assessment tool because the checklist component acted as a prompt for the peer-feedback group, allowing participants to gain familiarity with the steps of the procedure. The use of video recording allowed participants in the peer-feedback group to practice independently and still obtain relevant peer-feedback through video review. Although it is difficult to separate the educational effects of assessing the video footage from the other sources of feedback, the improvement in OSATS scores does suggest that using video footage for feedback was beneficial. This is consistent with other studies that have demonstrated an educational benefit of using video review (Soucisse et al., 2017; Vaughn et al., 2016). Reviewing video footage, however, takes time. Editing video footage for the purpose of assessment (Datta et al., 2006) is time intensive and may exclude critical issues; however, it does provide opportunities for focused feedback and reduces the time required for video review. Faculty did not use the recorded video for feedback purposes, as the faculty group was the control group for our study and as such, participants were taught in the traditional manner using contemporaneous feedback.
Improvement in technical skills was more pronounced in the bowel anastomosis study than in the skin excision and closure study. This result could have been influenced by the fidelity of the skin model, as participants found the thick porcine skin difficult to manipulate, especially when it dried out during the long duration of practice, thereby making it difficult to adequately demonstrate careful tissue handling. Additionally, the use of trained raters for video assessment, rather than in-situ real-time assessment, may have influenced our results. Nickel et al. (2016) noted that video assessment combined with procedural time may be more comparable to real-time scoring, due to the smoothing effect of reviewing the video at higher speed.

The duration of time required to perform each procedure was used as a form of formative feedback by participants in each group. We did not set specific time goals, however, participants often set personal time goals during the study. For the anastomosis study the overall improvement in time was statistically significant, but the clinical significance (2 minutes) is questionable, as the improvement between individual attempts was small. It is likely that the participants were pursuing quality improvement of the anastomosis over improvement in time to complete the task. In fact, our results demonstrate that for more complicated tasks time is a poor predictor of technical skill or the quality of the final product. Participants in Study 1 took longer to perform their second attempt than their first attempt, likely because they realized that poor technique was correlated with anastomotic failure during the leak test.

Our study assessed the quality of the final product – adequacy of bowel anastomosis and skin closure – which provided participants with visual feedback about the adequacy of their technique, information that was then used for self and peer-assessment. For example, leaks identified at the mesenteric border of the anastomosis became the focus of subsequent practice. Participants in the peer-feedback group felt that they did not possess the required technical
expertise to provide high quality feedback; however, they were able to provide highly relevant common-sense suggestions after seeing the quality of the final product. For example, after sutures tore through on a skin stress test, a suggestion was made to place the sutures wider on the next attempt.

Deliberate practice is dependent of provision of timely and accurate feedback. Participants in the peer-feedback group expressed concerns that their lack of expertise would put them at a disadvantage as compared to faculty-feedback group. This concern was highlighted in our post-intervention survey results. Some peer group participants, in fact, came to the investigators at the completion of the study asking for faculty feedback thinking that they received inadequate feedback from their peers. This subjective concern, however, was not supported by our objective results. Participants in both groups in both studies improved to the same degree regardless of the type of feedback that they received.

Despite the subjective fear about lack of expertise of participants in the peer-feedback group, students in this group were able to learn the skill without faculty support and felt able to provide useful feedback to their peers. This highlights that self-directed learning, which is heavily reliant on student’s internal motivation, can overcome this fear of inexperience, if appropriate structured support is provided. It is interesting to note that the second-year medical students were slightly more confident in their ability to provide feedback and learn independently. This confidence may have created a small power differential between peer-feedback participants on giving and receiving feedback, but baseline experience demonstrates that as a group they were comparable.

The results of the reflective activities survey demonstrate that some participants in each group, perceived that reflecting on one’s actions was a valuable, learnable skill that could be
used to improve another’s performance as well as his/her own. We did not actively promote the value of self-reflection in each study; however, we hypothesized that by integrating these reflective activities into our study, we could naturally increase the awareness of participants about the importance of self-reflection. Only a small subset of participants (9.1%) reported that self-review or reflection had benefits beyond direct performance improvement, but whether this translates to changes in behavior beyond the study intervention was outside the scope of this study.

Potential limitations in our research highlight the need for continued research in this area. The study participants were not trained to proficiency, so it is difficult to ascertain whether there are limitations in peer-feedback in acquisition of technical skills at higher levels of performance. We, however, believe that structured peer-feedback can be used in the early stages of learning allowing independent practice, prior to re-engagement with faculty for honing of skills prior to summative assessment. Second, the control and intervention groups were not spatially separated during the conduct of the study. It is possible that sharing of feedback suggestions may have occurred between participants in different groups. This does not necessarily detract from the results, as sharing of feedback suggestions would still be considered peer-based feedback. Third, the faculty-feedback group did not receive one-on-one feedback as the maximum faculty to student ratio was 1:8. This ratio was consistent with other laboratory skill-based activities; however, participants who were having more difficulties may have received more feedback than others. Participants in the peer-feedback group did receive one-on-one feedback with another peer group participant. It is possible that participants in the peer-feedback group received greater quantity of feedback; however, we were not able to quantify this in our study. Providing meaningful feedback is a time intensive process, the quality of feedback for both groups may
have suffered due to this constraint. In the faculty group, faculty members had to be shared amongst the participants. In the peer group wanted, peers wanted to return to their own practice. Whether the peer group viewed the assessment and feedback of another’s work as beneficial to their own is not known. Finally, the video recordings of practice sessions by the participants in the faculty-feedback participants were not utilized for feedback purposes in this group. Video recordings may have highlighted other areas for improvements not witnessed by the faculty, which could have been addressed at that stage.

We propose several areas of additional inquiry based on our findings. From the technical skill point of view, it would be valuable to see how this approach applies to advanced surgical skills, and whether more experienced participants are more comfortable giving and receiving structured peer-feedback. It would be valuable to examine the effect of incorporating peer-feedback into a post-graduate training curriculum to examine if it may be used to augment or replace faculty-feedback during acquisition of advanced surgical skills. Finally, given the importance of self-reflection as an essential life-long learning skill, it would be valuable to know whether the structured supports provided in this study were internalized by participants, and then applied outside of the study context.

**Conclusion**

If you provide students with the structural skills necessary to assess performance and provide feedback, then the peer-feedback provided is comparable to faculty feedback for the acquisition of both a basic and intermediate surgical skill. This means that students can be given the freedom to practice independently, without needing to have a faculty member present at all stages of learning a skill. Additionally, if self-review is built in as a natural part of the activity, some students will recognize
Chapter 4

Self-Regulated Learning Study

Abstract

Introduction.

The ability to develop evaluative judgement on one’s own or another’s performance is an essential non-technical skill, necessary to becoming a proficient physician. Unfortunately, little formal training is included in the current medical student and residency training curricula to develop these skills. Peer-feedback may help students develop evaluative judgement skills, but this is yet to be investigated within medical training. Using a skills-laboratory based study comparing faculty and self/peer group teaching methods, we conducted an in-depth analysis of the self/peer group to examine the progression of students’ evaluative judgement skills and the impact of these skills on their self-regulated learning (SRL).

Methods.

A qualitative thematic analysis of written peer-feedback and student survey responses was conducted. In addition to the two apriori hypotheses: progression of evaluative judgement and incorporation of evaluative judgement within students’ SRL schema, evidence of emergent themes was also sought. Surveys were conducted at baseline, at completion of training and after a minimum of three-month delay. Triangulation was performed using collected data and outcomes of technical study to improve confidence in the resultant themes.

Results.

Students developed increasingly sophisticated evaluative judgement skills with qualitative changes in the content and focus of the written peer-feedback. Students’ SRL was demonstrated by their improving technical performance, improving evaluative judgement of their
peer’s performance, self-identification of the benefit of these assessment skills to their own learning, and subsequent implementation of these skills in other contexts. Despite improved assessment skills; students did not believe that the feedback they provided to their peers was beneficial (despite the fact that it was equivalent to faculty feedback in terms of outcomes).

**Conclusion.**

Structured peer-feedback with appropriate support, can be used to augment traditional faculty led training through independent practice, whilst additionally developing students’ evaluative judgement and self-regulated learning. It is likely that implementation of these skills early within medical training would provide a strong foundation for life-long learning.
Introduction

An essential element of all medical practice is the ability to evaluate the quality of a skill to ensure that good clinical practice occurs. Traditionally, it was through personal exposure to multiple experiences of that skill that expert physicians developed the intuitive understanding of what quality looked like. Novice students and trainees are often reliant on the expertise of faculty to provide feedback on what quality performance of a particular skill should be. This method of teaching a clinical skill is dependent on the consistent availability of faculty, something that can be problematic in today’s busy clinical environment (Zevin et al., 2018). In addition, this traditional method of faculty driven feedback does little to develop the students’ ability to examine and evaluate their own learning. The ability to self-assess is an essential, non-technical professional competency that students must acquire to become independent and successful practitioners (Ryan, 2010). Evaluative judgement, defined as “the capability to make decisions about the quality of work of self and others” (Tai et al., 2018, p. 471) is a higher-level cognitive skill that must be taught and practiced in order for students to gain independence from faculty for their learning. Evaluative judgement has been found to be an essential element of life-long learning (Liu & Carless, 2006) and a valuable component for ongoing skill development.

Self-assessment and peer feedback are common methods to engage students in the assessment process and to build students’ evaluative judgement. Unfortunately, efforts to introduce peer or self-assessment in undergraduate medical education have been problematic. Student resistance to the assessment scoring process (Liu & Carless, 2006; Tai et al., 2018), perceived lack of student expertise about scoring (Liu & Carless, 2006), the negative motivational impact on students engaging in feedback (Hattie & Timperley, 2007), and the logistical requirements on faculty to incorporate these assessment and feedback activities (Liu &
Carless, 2006; Sadler, 2010) are some of the cited reasons for this problem. Concern about the grading and scoring process is common to much of this resistance. Many authors (Davis et al., 2006; Herrera-Almario et al., 2016; Hu et al., 2015; Porte et al., 2007) of studies on technical skills in surgery also raised similar concerns particularly towards the reliability of student scoring of technical performance as it compares with expert scoring. However, studies that examined performance improvement (Hu, Tiemann, & Brunt, 2013; Jensen et al., 2008; Zevin, 2012) through formative assessment mechanisms, such as self-assessment or peer review, did demonstrate a benefit. The metacognitive understanding that develops through the students’ participation in the formative peer and self-assessment studies is yet to be sufficiently explored.

Within formative assessment, it is through the communication process that students learn how to assess a task (Tai et al., 2018). As such, the term feedback will be used in this manuscript to highlight the communication process that is used to discuss the performance of a task against a defined standard (Liu & Carless, 2006). In peer-feedback, where neither student claims expertise, there is an absence of the power dynamics found in traditional expert based coaching models. The feedback can be described as a learning dialogue: an active two-way dialogue form of feedback that enables free sharing of ideas and concerns between peers (Askew, 2000). A learning dialogue allows “feedback and reflection [to] become entwined, enabling the learner to review their learning in its context and relate [it] to previous experiences and understandings” (Askew, 2000, p. 13). This sharing of ideas and concerns reduces the barriers to students engaging in the evaluation activity. Its potential to support subsequent learning has already been demonstrated. In an engineering context, peer-feedback in a technical design project was found to both improve the technical quality of the product, as well as provide students with insights into their own evaluative judgement and learning (Nicol et al., 2014). The potential for such
peer-feedback within the undergraduate medical education curriculum to improve students’ evaluative judgement (Tai et al., 2018) is present but has yet to be fully explored.

The ability to develop students’ evaluative judgement in undergraduate medical education can allow them to become independent life-long learners who are also able to more effectively monitor their own practice (Mann, 2016). The internalization of these evaluative skills within the students’ learning armamentarium can be described through the process of self-regulated learning. Self-regulated learning theory posits that self-improvement is based on a combination of self-motivated beliefs and behavioral learning strategies that are iteratively applied in a self-regulated way (Zimmerman, 2002). Specifically, Zimmerman (2002) describes three phases within the cyclical process of self-improvement: a forethought phase, a performance phase, and a self-reflection phase. Each phase has its own inherent tasks and motivational elements that build upon and are reinforced by the preceding phase.

To understand self-regulated learning in the medical education context, a spectrum of approaches have previously been used, from qualitative semi-structured interviews, structured interviews (microanalysis), through to quantitative surveys (Cho, Marjadi, Langendyk, & Hu, 2017). Foundational to each approach is a structure built upon Zimmerman’s three phases. Each approach has its own benefit, and as such no one method has been shown to be superior than another. Therefore, for this research we chose to return to the original theory and use the three phases as our structural framework, rather than rely on another existing method.

The purpose of our research study was to examine the progression of novice medical students’ evaluative judgements of a technical skill without ongoing faculty guidance and the impact the development of evaluative judgement had on their self-regulated learning. Our study had two apriori hypotheses: (1) students will develop evaluative judgement skills without faculty
input during a technical skills study, and (2) students will incorporate the newly acquired evaluative judgement skills into their self-regulated learning schemas.

Methods

For this study, we retrospectively analyzed previously collected, but not reported, self-assessment, personal review and peer-feedback data for participants in the peer-feedback group of two randomized controlled trials (RCT) (Sheahan, Reznick, Klinger, Flynn, & Zevin, 2018). In these RCT’s, first and second year medical students were taught a skin lesion excision and suture closure or a handsewn bowel anastomosis, respectively. Students performed at least five attempts of the skill over four days whilst being provided ongoing feedback, based on randomization to either the traditional manner from faculty, or using peer-feedback with structured support (Sheahan et al., 2018). The feedback used by the peer-feedback group consisted of a combination of self-assessment, personal review and peer-feedback (learning dialogue). Students used the structured support to informally self-assess their own work. This structured support consisted of pre-training in performance assessment with an objective structured assessment of technical skills tool (Martin et al., 1997) (OSATS) using video frame of reference training. The OSATS is a validated tool that assesses performance using a task-specific checklist and global rating scale. Students also reviewed the overall quality of their finished product after each attempt by observing the bowel or skin closure undergo a functional stress test. Finally, a peer reviewer independently assessed the videoed attempt with an OSATS tool and used the OSATS score as a guide to discuss the performance with the individual via a learning dialogue form of feedback. Prior to beginning our research ethics approval was obtained from the university ethics board (Queen’s University IRB no: 6020826).

Written feedback.
Students in the peer-feedback group were requested to write down the feedback they provided during the peer-feedback process to facilitate review by the recipient of the feedback, and for collation and subsequent analysis by investigators.

**Surveys.**

Three online surveys were administered to capture the opinions of the students on the technical skills, the assessment and feedback activities, and the perceived impact of these activities on students’ self-regulated learning. There was a baseline survey at the commencement of the study, a post study survey at the completion of the study and a delayed survey after a minimum three months post study completion. Questions differed between each survey and consisted of a mixture of quantitative (1-5) Likert-type scale questions and short answer response questions.

**Baseline survey.**

The baseline survey collected demographics data and students’ reported experiences with suturing, skin excision and bowel anastomosis (Appendix B). Students’ expectations of their own performance, and their self-perceived ability to learn skin excision and bowel anastomosis were also sought.

**Post study survey.**

Post-study survey collected students’ views on the role of feedback, methods of delivering feedback, and the implementation of peer-feedback taught in the study (Appendix C). Student views on the intended use of the newly acquired assessment skills were also collected.

**Delayed study survey.**

Three to 12 months after study completion, students were asked to reflect on their self-regulated learning during the study (Appendix D). Specifically, their attitudes toward the
assessment tools and the feedback skills learned for the study, the effect of these skills on their own and other’s learning, and finally, the subsequent impact of these new skills on their own attitudes towards learning technical skills. This delayed survey re-evaluated many of the themes covered in the earlier surveys. This provided two benefits, consistent responses between surveys would further reinforce the validity of the students’ responses, whereas responses that altered over time, would suggest students developed new insights into the assessment process.

Analysis

Agreement with the Likert-type scale questions were reported as percentages. To examine the developing evaluative judgement and self-regulated learning of students a thematic analysis (Creswell, 2013) of the short-answer survey responses and written peer-feedback was performed. Massey's (2011) interpretation of thematic analysis defines three levels of thematic data: articulated data – the questions directly asked, attributional data – any apriori hypotheses, and finally emergent data – themes that arise from the data itself. To address the first hypothesis of developing evaluative judgement, we examined the written peer-feedback for qualitative differences in the content written over the course of the study. To address the second hypothesis of the positive impact on students self-regulated learning, we used triangulation of the results of the technical study, written peer-feedback and the survey results, as a means of developing confidence in the students self-regulated learning.

Results

Technical performance

The technical performance by the faculty and peer-feedback groups improved over the five attempts for both skills: skin excision and suture closure, and handsewn bowel anastomosis.
The performance improvement was similar between the two alternative feedback groups (faculty vs peer/self) across all assessment criteria (Sheahan et al., 2018).

**Written peer-feedback**

Ninety-two completed OSATS tools were available for analysis of emergent themes. Five themes emerged from the analysis of written peer feedback: (1) non-specific OSATS feedback, (2) instructive OSATS feedback, (3) quality-related feedback, (4) higher-order feedback and (5) students’ sentiments toward the feedback process.

Non-specific OSATS feedback (Theme 1) referenced either the checklist or global rating scale elements of OSATS tool, but did not provide any instructive feedback on how to address the identified problem. Examples included “longer incision”, “tighter sutures” and “reduce unnecessary motions”. Feedback of this type was evident across all practice attempts, however the checklist component of the OSATS was used to provide feedback early on in the study, whereas comments involving the global rating scale (GRS) component of the OSATS showed a biphasic distribution due to an initial focus on procedural duration, whilst later attempts focusing on more general quality elements (Table 6).

**Table 6. Frequency of theme 1: non-specific OSATS feedback**

<table>
<thead>
<tr>
<th>Attempt</th>
<th>OSATS</th>
<th>Checklist</th>
<th>GRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.52</td>
<td>86%</td>
<td>21%</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
<td>87%</td>
<td>13%</td>
</tr>
<tr>
<td>3</td>
<td>0.51</td>
<td>77%</td>
<td>24%</td>
</tr>
<tr>
<td>4</td>
<td>0.31</td>
<td>68%</td>
<td>44%</td>
</tr>
<tr>
<td>5</td>
<td>0.11</td>
<td>42%</td>
<td>58%</td>
</tr>
<tr>
<td>6&amp;7</td>
<td>0.21</td>
<td>10%</td>
<td>90%</td>
</tr>
</tbody>
</table>

Instructive OSATS feedback (Theme 2), again referenced the OSATS tool, but now included specific instructions on how to improve a student’s performance. For example: “suture
ends cut shorter”, “use forceps to grasp needle”, “reduce tension of sutures”, “deeper cut”, and “try to keep the length of your stays longer as they were getting in your way” (italics added).

Instructive feedback accounted for 30% of all written feedback and was distributed equally throughout all practice attempts.

Quality-related feedback (Theme 3) focused on improving the overall quality and efficiency of the performance. Although grounded in the global rating scale elements of the OSATS, responses found in this theme were not limited to it. Quality-related feedback consisted of two separate elements: a speed element, for example, “be faster 25 mins”, “pace could be faster”, and an indirect quality element. The indirect quality element (e.g. “more aesthetically pleasing”, “better ergonomics”, “better planning”) was only present from the third attempt onwards and increased in frequency as participants continued to practice. Feedback related to speed was uniform throughout all practice sessions (Table 7).

Table 7. Frequency of theme 2: quality-related feedback

<table>
<thead>
<tr>
<th>Attempt</th>
<th>Theme 3: OSATS</th>
<th>Speed</th>
<th>Indirect Quality Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.09</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>2</td>
<td>0.05</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>3</td>
<td>0.11</td>
<td>27%</td>
<td>73%</td>
</tr>
<tr>
<td>4</td>
<td>0.29</td>
<td>49%</td>
<td>51%</td>
</tr>
<tr>
<td>5</td>
<td>0.29</td>
<td>55%</td>
<td>45%</td>
</tr>
<tr>
<td>6 &amp; 7</td>
<td>0.31</td>
<td>39%</td>
<td>61%</td>
</tr>
</tbody>
</table>

Higher-order feedback (Theme 4) was related to higher level thinking and better forward planning. Examples include, “more forward planning (think of a sequential plan) to be more efficient”, and “planning of suture positioning ahead”. This type of feedback was rare, occurring only three times in the study and all on the fourth attempt.
Students’ attitudes toward the feedback process was the fifth and final theme identified in the written feedback. This type of feedback was evident throughout all practice attempts, although two contrasting attitudes were noted. Students who hesitated to provide feedback wrote: “improve everything” (after the first attempt), and “improve overall technique” (after the seventh attempt). The first “everything” suggests that the student felt overwhelmed due to inexperience and did not know where to begin making suggestions for improvement. The final generic “overall technique” comment suggests the that performance exceeded student’s expertise and the student now felt unable to make any specific suggestions. This conclusion was corroborated by subsequent delayed survey responses suggesting ‘I found I was running out of suggestions to give to my peer and vice versa’ and that OSATS scores by reviewers were consistent with the performance attempt e.g. OSATS score of 34% on the first attempt, and 85% on the seventh.

Students’ positive engagement with peer-feedback was demonstrated with the use of suggestive terminology, such as: “*try* vertical mattress sutures”, “*try* holding needle 2/3 1/3 from tip”, “*you could* apply more countertraction on exit to tug on tissue less”, and “*you could* take bigger bites” (italics added). This use of suggestive terminology occurred with increasing frequency as the study progressed and technical skills of students improved. Confidence was also evident when personal pronouns were used: “*I think* your sutures are long and the tails are getting in your way”, and “*I think* suture tails getting in your way as sutures look too close together” (italics added). In addition to demonstrating confidence in a peer reviewer’s own skills, these findings also highlight the collaborative, dialogue approach to feedback, whereby personal suggestions made by the reviewer can be dismissed by the recipient if felt to be unhelpful. General positive reinforcement was also provided: “good planning”, “*more consistent throughout – improves as I went*”. Positive comments were provided in isolation from other comments, but
also used as a feedback framing technique, for example, “your needle entry is very smooth, good job, you could apply more countertraction on exit to tug on tissue less”.

**Surveys**

**Baseline survey**

The baseline survey was completed by 87% of study participants. Students in both studies were familiar with suturing but were essentially naïve to the specific technical task (skin lesion excision and suture closure or the bowel anastomosis). Students rated themselves low (2.6 out of 10) on their baseline ability to perform the task but had moderate confidence in their ability to learn the task with practice (6.4 out of 10). Sixty-eight percent of students did not feel they had enough time to practice practical skills in medical school and 92% of students desired more time to do so.

**Post study survey**

Students had three main comments about the assessment and feedback process during the studies. They reported positive views about the study, the methods used to teach the technical skills, and the benefits of practice of the newly acquired skills. All but one student (1.6%) wanted to be randomized to the faculty group, primarily due to impressions about the quality of the peer-feedback available. “In my final attempt, I received expert feedback and it was very, very useful in perfecting my technique”. The one student felt that “it seemed like the peer group would get more feedback. At my very beginner stage, I think more feedback would have been useful for improving my technique instead of less expert feedback”.

This concern about the quality of peer-feedback was based on both the peer group’s inexperience in providing feedback and the uncertainty about using the assessment tools. For example, “some peer comments were not highly critical and seemed nervous to give any
negative feedback”, “I felt as though in the peer group, how we were supposed to assess one another was unclear and therefore feedback was very different from day to day” and it “took a few tries to get comfortable, but mostly because of our unfamiliarity with the tools and procedure at this stage of training” and “waiting for peers to review made it more time consuming”.

Despite wanting faculty feedback, most students (75%) reported that they were comfortable learning the skill with feedback from their peers. Most students (79%) found the peer-feedback process easy to complete, and 65% felt they were able to provide specific rather than general feedback, with the second-year students being more comfortable applying these skills than the first-year students. Sixty-five percent of first-year students felt they still lacked the skills necessary to assess tasks other than the skin excision and suture closure taught during the study, but 85% planned to use these same skills in the future for their own learning.

**Delayed survey**

Twenty-six out of 61 students (42%) completed the survey. Three themes emerged from the analysis of delayed survey responses: freedom to experiment, appropriation, and reflections on the feedback.

Freedom to experiment was based on the students’ attitudes and motivations, and the students’ perceptions of the study design itself. Most students (89%) felt that the study design created a permissive environment in which to experiment, but this experimentation extended only to the task itself: “I personally would have done better with more than 5-8 trials. I think my experimenting of suture distance would have eventually worked itself out. Because feedback from others who were more successful did not translate quick enough”. Most students (76%) reported that they experimented over 50% of the time with how they performed the task because they felt they had technical skills and knowledge required to experiment. Students did not seek
ways of experimenting with other aspects of the learning process. For example, when asked why they did not ask to re-watch the demonstration video, students reported: “did not even think about this as an option, but also found the peer feedback and support was sufficient”.

Appropriation highlights where students have adopted the assessment and feedback skills that were acquired during the study. Broadly, students felt that this study changed the way they consider their own or others work in a small to moderate way.

Yes, the OSATS helped immensely, especially the front part of the sheet listing specific items/gestures to attain proficiency in. Having previous feedback forms available to refer to on subsequent attempts was helpful in refreshing memory/tracking progress/patterns of e.g. bad habits. Reviewing my own videos helped me realize areas of inefficiency (e.g. placement of tools) that I would not have otherwise paid attention to.

More specifically, most students (75%) felt that the purpose of the assessment process was to enable feedback, and based on these skills, had or intended to apply them in the future. Nearly half of the students (47%) had already used an OSATS or other assessment tool to assess performance in another context.

The final theme was that of reflections on the feedback process. In a similar fashion to that of the completion survey, students felt that peer inexperience was a significant impediment to the acquisition of technical skills. “I think it was helpful if the other person had a successful technique, but without either individual really know[ing] what was ‘right’ it was difficult to provide constructive feedback”. Fifty-three percent of students did not believe that the feedback they personally provided was helpful to their colleagues, nor did they feel that they had the expertise or knowledge to provide feedback: “I found I was running out of suggestions to give to
my peer and vice versa”. Only 43% of students found receiving feedback from peers helpful to their performance. Yet despite this perceived inadequacy of the quality of the students’ own feedback, 75% of students found the process beneficial for their own learning:

Analyzing my peer’s video with the mindset of constructive criticism helped me learn how to break down the large task into smaller items, and ultimately, I believe, helped me concentrate better on specific actionable items when I was suturing myself,

and “structured feedback was useful and guided my own learning. ‘What are metrics we measure and how can I focus on each of those on subsequent iterations?’ Learning how to give feedback was very educational and guided my own learning”.

Discussion

We used a thematic analysis to explore students developing evaluative judgement and the impact it had on their self-regulated learning. Our results demonstrate that through structured peer-feedback, students developed more nuanced evaluative judgement skills. The results also highlight that these developing judgment skills did have a meaningful short and medium-term impact on students self-regulated learning. The short-term impact on students self-regulated learning was improved technical performance during the study, whilst the medium-term impact was the planned or implemented use of these evaluative judgement skills in contexts outside of the study. Aside from the validation of the apriori hypotheses of this analysis, there are two other significant conclusions from this research. Firstly, non-technical skills such as evaluative judgement can be successfully implemented during technical skills training. Secondly, structured peer-feedback, as a supportive addition to faculty feedback, enables students to take more ownership over their own learning, and consequently provides evidence to encourage the use of
guided independent practice within a technical skills training curriculum. It is therefore plausible that early exposure in medical training to the evaluative judgement training used in structured peer-feedback would be highly valuable to students as they seek to establish patterns of life-long learning. Before we discuss the implications of these conclusions, it is necessary for us to go back and examine the justification for why these conclusions have been made.

The thematic analysis conducted in this research explored students’ developing evaluative judgement and self-regulated learning whilst being conscious of any emerging themes that might appear. The first hypothesis postulated that students’ evaluative judgements would improve through the practice with structured peer-feedback during the technical skills study. Evaluative judgement of a technical skill requires students to examine “inward on one’s work, but also outwards to others’ performance, about the two in comparison to each other, and standards . . . while also transcending the immediate task by developing a student’s judgement of quality that can be further refined and evoked in similar tasks” (Tai et al., 2018, p. 475). We were able to demonstrate progression in evaluative judgement in our study on two levels: an outcome level and a cognitive level.

On the outcome level, students’ technical performances improved throughout the study, therefore students’ evaluative judgements must have improved. This improvement in performance was beyond that caused by repetition, because the amount of improvement was comparable to students receiving feedback from trained faculty familiar with the procedure.

On the cognitive level, the written feedback demonstrates the students’ evolving thoughts about what defines a “good” performance, and how to share that information with their peers. The feedback provided throughout was informative, even if the correct terminology was not known or used. There was a general shift away from checklist items toward global quality
elements, such as time, efficiency and planning in the later stages of the study. Students recognized that other factors outside of the OSATS tool could have a significant impact on the overall performance of the task, such as ergonomics and the results of the functional test (anastomotic burst pressure and wound strength). The language used by students also progressed throughout the study for example, personalization (e.g. “try” or “consider” or “I”) demonstrated increasing confidence in their attempts of providing feedback in a dialogic manner. Both members of the learning dialogue journeyed together as they learned the task, the assessment and the feedback process. As one student described:

There of course were things on the checklist people could improve upon, but I valued more the opportunity to receive (or give) comments from (or to) other students about the process/technical skills in general.

To explore the second hypothesis, the impact of developing evaluative judgement on students’ self-regulated learning, consistent features in the forethought, performance and self-reflection phases of the students’ self-regulated learning were identified using triangulation of available data. The forethought phase consists of a learner motivation element combined with a task analysis element. Student motivation in our study was clearly mixed. Motivation toward the technical study was overwhelmingly positive. Students volunteered for the study because of the opportunity to learn and practice the technical skill. The survey responses demonstrated that students were confident they could learn the skill and this confidence flowed into the independent experimentation students reported throughout the study. Students, however, had what might be termed “negative motivation” towards the assessment and feedback aspects of the study. As described in the expectancy-value theory of motivation (Eccles & Wigfield, 2002), students were externally driven by the expectancy and costs of peer-feedback. Students focused
extensively on their or their peer’s lack of expertise and the potential cost this had on their progress. Hence, almost all students wanted to receive some form of faculty feedback. This expectancy of receiving poor quality feedback was mitigated during the study by students increasing familiarity and comfort with the technical task, the assessment and feedback process. The opportunity to have multiple practice sessions generated information used in the planning phase of the subsequent attempt.

The performance phase of self-regulated learning focusses on the tasks used by students during the conduct of the activity, as well as the self-observation that occurs during this phase. Self-observation by students during the study focused on the perceived lack of expertise in themselves and their peers to conduct the assessment and feedback tasks. This perception was incongruent with their objective improvement in performance, and the self-reported benefit that peer-feedback was to their own performance. Although only 43% of students found the information provided by the peer-feedback helpful (unhelpful (19%) or neutral (38%)), the active process of engaging in an evaluative judgement through assessing and discussing the performance helped inform the students own performance (75%). Finally, this perceived lack of expertise also impacted the ongoing progress of their skill acquisition: “I enjoyed it but felt that I plateaued in my improvement . . . only continue to improve by doing repeated practice. I am unsure how I would alter my approach to improve my technique”.

The final phase of self-regulated learning is that of self-reflection on the completed activity. Students’ felt that these new assessment skills were valuable, initially reporting their intent to use these evaluative judgement skills on completion of the study. This intent was subsequently confirmed in the delayed survey, where students after the study reported utilizing these skills to assess technical performance of a different task, or still planned to utilize these
skills in the future. This student appropriation highlights the higher order self-regulated level of feedback described by Hattie and Timperley (2007). Similar to other studies (Nicol et al., 2014), students found the process of giving peer-feedback beneficial to their own learning – something only identified by students on reflection after the study. One of the benefits of using a learning dialogue with a peer over pure self-assessment is that “Learning about notions of quality through assessing the work of peers may reduce the effect of cognitive biases directed towards the self when undertaking self-assessment” (Tai et al., 2018, p. 475). However, there was still a strong discord in the self-efficacy of the students’ feedback skills, as has been highlighted above, students found the evaluative judgement process of peer-assessment and discussion more valuable than the receiving of feedback from peers. Yet, despite this recognized benefit from the feedback process, and demonstrated performance improvement students continued to feel that their skills and knowledge were inadequate.

The thematic analysis was also sensitive to any emergent themes that arose in the data. Across each data source a common theme kept occurring: the novice students felt they did not have the skills and knowledge necessary to adequately assess or provide peer-feedback. This perception was despite demonstrable evidence to the contrary. This theme was found in the completion and delayed survey as well as within the written feedback. This perceived lack of expertise highlights two important principles. Firstly, the expertise of faculty is an essential element to technical skills training; however, adjuncts such as structured peer-feedback can shift some of the onus of learning from faculty onto the students themselves, thereby enabling students to practice the skill independently. Secondly, faculty must nurture the students’ developing evaluative judgement. This support includes reinforcement about the students
improved performance and perhaps more importantly, provide feedback about the students developing evaluative judgement skills.

**Limitations**

Pre-existing experience of students in providing peer-feedback and assessment were not measured at the commencement of the study, which did not allow us to define feedback expertise trajectories. The peer-feedback training focused on understanding and using the OSATS tool but did not contain any formal discussion about how to deliver feedback to maximize learning. This may have improved student comfort towards the feedback process and altered their perceptions towards the feedback process, in a similar fashion to the assessment process. Technical issues – storage space of recording devices, prevented the recording of the verbal component of the assessment and feedback process, which would have provided richer information to analyze.

Finally, the technical study used a deliberate practice approach to the task acquisition, however this same approach was not applied to the feedback and assessment process. Students were not provided any feedback from faculty about their peer-feedback, and as such remained ignorant of the quality feedback, and increasing expertise they were demonstrating in this process. This may have been a factor in why students’ assessment of their own feedback skills was so poor. Students were heavily invested in improving their technical performance, but often felt frustrated by the amount of time required to complete the assessment process. Hence students chose not to invest any time in reviewing and reflecting on the assessment and feedback process itself, as this was not mandated by the study.

**Further research**

One of the challenges of peer-feedback as it was expressed in this study was that it was heavily focused on corrective information, as students were helping align the performance with
the standard as defined by the assessment criteria. It would be highly informative if students were instructed to use a more inquisitorial approach, such as: what length of suture tail would be most useful? or what could be done to increase the tightness of the sutures? This approach may produce unorthodox less prescriptive performances but may enhance students’ evaluative judgement. It would be beneficial to examine the translation of these peer-feedback skills in a self-assessment context. Although the assumption that these skills will be translated is reasonable, especially given appropriation of these skills as part of their self-regulated learning, this has not been formally explored. How students apply these new evaluative judgment skills onto other technical, or even non-technical skills has yet to be examined. As novice learners, unfamiliar with technical skills beyond basic interrupted instrument suturing there was an expectation that students would require explicit assessment training on the chosen OSATS tool.

It may be possible that more experienced students or residents, could develop their own assessment criteria, through a faculty supported consensus process, which may further boost their evaluative judgement skills. Additionally, the inclusion of formal feedback and assessment training early within medical training would likely be beneficial for both students self-regulated learning and non-technical skills acquisition, but this also requires further exploration.

**Implications**

Having now explored the rationale for the conclusions drawn from this study, it is valuable to examine the implications of this research to the medical education community. The importance of these conclusions is that they provide mechanisms by which technical skills training can be provided to meet some of challenges currently facing the medical education environment (Bolster & Rourke, 2015; Jones, Higgs, de Angelis, & Prideaux, 2001; Mamede, Schmidt, & Rikers, 2007; Mann, 2016; van de Wiel & Van den Bossche, 2013). Part of the
solution to these challenges lies in building students’ metacognitive skills and knowledge, because modern educational theory seeks to enhance the independent learning of students by encouraging and developing the inherent learning skills of the students via activities that can promote life-long learning (Brydges & Butler, 2012; Garrison, 1991; Ryan, 2010). Efforts to build students’ non-technical skills and professional competencies such as self-assessment, early in their career, will assist them in becoming successful practitioners (Ryan, 2010). However, adding non-technical skill training into already stretched curricula is difficult (Jones et al., 2001), especially considering the increasing clinical demands placed upon faculty. Ancillary activities, such as guided independent practice using structured peer-assessment, may be able to integrate these non-technical skills into existing technical skills training.
Chapter 5

General Discussion

The overall aim of this thesis was to examine the effect of introducing assessment training to novice medical students’ technical performance and examining the impact of this training on their metacognitive understanding. Studies examining non-expert forms of feedback have primarily focused on technical performance, and yet little is understood about the impact this involvement has on the participants themselves. Self-regulated learning theory was used to assess the student’s metacognitive understanding as it provides a method of incorporating the external aspects of assessment with the higher processes of personal motivation, planning and critical reflection. Results from student participation in this technical skills study demonstrate that when provided appropriate assessment training and feedback support with an accessible assessment tool such as the OSATS that skill acquisition of novice students for a simple and intermediate surgical skill is comparable to receiving traditional faculty feedback. What is more crucial is that in addition to learning these technical skills, peer-feedback students also improved their evaluative judgment through assessment and feedback skills, and that the students felt that these new skills were valuable enough to add to their self-regulated learning skillset. Each chapter has already discussed the main findings of each study; however, I will highlight the main points below as well as discuss the limitations and implications of this research.

Technical skills study

The technical skill study (Chapter 3) used a non-inferiority methodology to assess the skill acquisition of novice medical students learning a skill using structured peer-feedback against traditional faculty feedback. Two trials, one a simple procedure: skin lesion excision and suture closure, and the other an intermediate procedure: handsewn bowel anastomosis, assessed
the performance of students over five attempts made in four days. Assessment of the task was by blinded-expert video OSATS assessments, time to completion and functional stress test results. For each assessment domain, and across both trials, performance in the peer-feedback group was not-inferior to the faculty-feedback group. This study demonstrates that with appropriate training novice learners can practice technical skills with peer feedback and without faculty feedback with ongoing improvement in performance.

Surprisingly, the rate of performance improvement varied across each criteria and task. This variation may be reflective of the limitations of the simulated tissue and/or the means of assessment. In addition to the limitations of the simulated task, logistical factors resulted in both faculty and peer-feedback groups sharing the same laboratory which may have cause some cross-pollination due to overheard feedback. Students found the process of video review and assessment laborious, and this along with individual preference may have influenced the extent of the learning dialogue that occurred in the feedback process. Finally, this study examined the early stages of skill acquisition with students only completing five attempts. More attempts would be required to examine whether the degree of improvement remains comparable, or whether at higher levels of performance faculty feedback exceeds peer-feedback.

Having now confirmed that structured peer-feedback is comparable across basic and intermediate surgical tasks, it would beneficial to examine whether this effect holds during training to proficiency and the effect of this approach in different contexts. Whether more experienced learners with greater skill and content knowledge, for example advanced-suturing skills, are more comfortable with this peer-based approach, and whether these results are also demonstrable with advanced surgical skills. Finally, research is required in how best to implement structured peer-feedback into a training curriculum.
Self-regulated learning study

The self-regulated learning study (Chapter 4) examined the non-technical and metacognitive aspects of the learners involved in the technical skill study. This examination was performed using a thematic approach to the qualitative data garnered through multiple surveys collected at the commencement, completion and after a three-month delay after the technical study. These surveys were analyzed independently for emergent themes prior to collectively for evidence of self-regulated learning theory implementation. This qualitative approach was also used to examine the written feedback provided by participants in the peer-feedback group.

Assessment of this written feedback revealed that students developed an increasingly complex evaluative judgement. As the students’ familiarity with the task and assessment increased, their expectations of what a ‘good’ performance increased as well, such that global quality aspects of performance such as efficiency, became the focus of their feedback. From the surveys, it became evident that despite evidence of improved performance, and a recognition of the benefit of peer-feedback on their own performance, students still felt that they lacked the skill and expertise to adequately provide feedback. Despite this apparent dissonance, students chose to adopt these skills into their self-regulated learning schema by planning or implementing these new evaluative judgement skills into contexts outside of the study environment.

The limitations of this study as highlighted earlier are based on the necessary precedence of the technical study. As such, structural support training focused primarily on the assessment component of feedback, rather than the delivery and dialogic process. Only written peer-feedback was collated and so opportunities to gather other feedback information such as the extent of dialogic process was lost. Students’ felt that their assessment and feedback skills were inadequate, and this potentially was due to the lack of faculty reinforcement of these developing
skills. The technical study used deliberate practice principles to practice the task, but this was not applied to assessment and feedback elements.

Given the success of the technical study, the results from the evaluative judgement study become more important. If novice students are provided with the skills necessary to appropriately assess a technical task, then, in addition to improving technically, they also develop evaluative judgement skills, which are incorporated into student’s self-regulated learning armamentarium. The next step for this research must be to study the influence that faculty support of these assessment and feedback skills has on participants assessment skills, self-regulated learning and self-efficacy. This could be achieved by comparing the performance and student perceptions to the same tasks with feedback coaching by faculty. Following from that, it would be valuable to examine the longer-term impact on student’s self-regulated learning using a longitudinal study to assess whether these skills taught early in their training are utilized for subsequent skill acquisition.

A framework by which quality of feedback can be measured is another finding of this research. In this study we looked at changes in written feedback to measure changes to student’s evaluative judgement. This approach relied on emergent themes becoming evident during the qualitative analysis. These identified feedback themes (non-specific, instructive, quality, higher-order, and attitudes) can now be used as the basis to measure quality of feedback as it pertains to evaluative judgement. Future studies could examine each individual aspect of the evaluative judgement: the assessment phase, the interpretation phase or the delivery process. Even within each phase quality could be examined either for progression, like our study, or to compare different educational interventions. Additionally, students’ perceptions of the quality of the feedback provided or received could also be used as a means of measuring quality, although as
we’ve identified student perceptions on the quality of feedback may differ from the quality provided.

**Conclusion**

This research sought to integrate the goals of improving students’ technical ability and improving their non-technical skills in a single activity. To achieve this, structured peer-based assessment and feedback was used to both provide the means of improving technical performance and opportunity to practice these skills. This research has demonstrated that structured peer-feedback can be used as a successful adjunct to faculty-led feedback, allowing students to practice independently without concern of technical regression. In addition to the technical gains made through the independent practice, structured peer-feedback also improves the learners’ evaluative judgement and self-regulated learning, both essential non-technical skills for physicians in today’s medical environment. This research also demonstrates how reflective tasks can be effectively implemented into a technical skill training curriculum. This research has also highlighted how student perceptions on their experience as assessors can impact the interpretation of their subsequent performance.
References


92


Appendix A
Reflective prompts built into OSATS

<table>
<thead>
<tr>
<th>Global Assessment</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respect for tissue</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why did you give this score?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Time and motion</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Instrument handling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Knowledge of Instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Flow of Operation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Knowledge of specific procedure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why?</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Total Score: checklist + Global Scale:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other points to consider reviewing:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ergonomics of body position: correct height, neck position</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Try focusing on one element</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Set a time goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you compared your performance with a video?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Goals for next attempt:
Appendix B
Baseline survey questions

<table>
<thead>
<tr>
<th>Age</th>
<th>Open response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Open response</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How many times have you done any suturing?</th>
<th>Open response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Have you seen or done the procedure before?</td>
<td>No</td>
</tr>
<tr>
<td>How would you rate your current ability to do the procedure?</td>
<td>0-10</td>
</tr>
<tr>
<td>How would you rate your confidence to learn to do the procedure</td>
<td>0-10</td>
</tr>
</tbody>
</table>
### Appendix C

#### Post study survey questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Did you think this study/education was successful in teaching you to do the skill?</td>
<td>No</td>
</tr>
<tr>
<td>How comfortable were you learning to the procedure without an expert to guide you?</td>
<td>Very uncomfortable</td>
</tr>
<tr>
<td>Did the teaching session accommodate your learning style?</td>
<td>No</td>
</tr>
<tr>
<td>Biggest strengths of the teaching session/study</td>
<td>Open response</td>
</tr>
<tr>
<td>Biggest weaknesses of the teaching session/study</td>
<td>Open response</td>
</tr>
<tr>
<td>What did you learn from this activity/study?</td>
<td>Open response</td>
</tr>
<tr>
<td>Any other comments?</td>
<td>Open response</td>
</tr>
<tr>
<td>Which group: peer or expert, did you want to be in for learning how to do the procedure? Why?</td>
<td>Open response</td>
</tr>
<tr>
<td>Would you have preferred to have some expert feedback in addition to the peer-feedback?</td>
<td>No Attempts</td>
</tr>
<tr>
<td>If you wanted expert feedback in addition to peer-feedback, when would you like it? - choose all that apply</td>
<td>All the way through</td>
</tr>
<tr>
<td>How comfortable were you reviewing/scoring your work with a peer?</td>
<td>Very uncomfortable</td>
</tr>
<tr>
<td>Throughout the study, when reviewing someone else performance did you feel you were able to provide more specific feedback in reviewing their work?</td>
<td></td>
</tr>
<tr>
<td>Question</td>
<td>Not at all</td>
</tr>
<tr>
<td>-------------------------------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td>Throughout the study, when reviewing someone else performance, were you more aware of things you liked or disliked in their performance?</td>
<td></td>
</tr>
<tr>
<td>Based on the skills learned in this study (reviewing video, using OSATS), how confident do you feel in assessing other procedures?</td>
<td></td>
</tr>
<tr>
<td>Based on the skills learned in this study (reviewing video, using OSATS), Do you plan to/or think you will use these skills for your own future learning?</td>
<td></td>
</tr>
</tbody>
</table>
## Appendix D
### Delayed study survey questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Peer</th>
<th>Faculty</th>
<th>Both</th>
<th>No</th>
<th>Not really</th>
<th>Neutral</th>
<th>Somewhat</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which group were you in?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Which study were you involved in?</td>
<td>Skin</td>
<td>Bowel</td>
<td>Both</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you feel you had the knowledge/skills/expertise to improvise or try different ways of performing the task?</td>
<td>No</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you feel the study gave you the freedom (did you feel you were allowed to) to improvise or try different ways of performing the task?</td>
<td>No, I had to follow the prescribe method</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes, I had complete freedom</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you experiment with how to perform the task?</td>
<td>Never</td>
<td>Some attempts</td>
<td>Half the attempts</td>
<td>Most attempts</td>
<td>Always</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I noticed that none of you asked to re-examined the expert video during study, why do you think that was? eg. Do you think that the more immediate feedback was preferable?</td>
<td>Open response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any other comments about the study?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Open response</td>
</tr>
<tr>
<td>How has your involvement in this study impacted the way you consider your work and the work of others?</td>
<td>No Attempts</td>
<td>Some attempts</td>
<td>Half the attempts</td>
<td>Most attempts</td>
<td>All attempts</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How have you, or how do you intend to be more analytical when you assess your/others performance?</td>
<td>Global way, that was good/bad</td>
<td>Some attempts</td>
<td>Half the attempts</td>
<td>Most attempts</td>
<td>Structured, deliberate way, I liked/not like this specific element</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Have you applied the OSATS type of structure (checklist of task + global rating scales 1-5)?</td>
<td>Never</td>
<td>Rarely</td>
<td>Half the time</td>
<td>Most of the time</td>
<td>Every time I do a procedure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What do you consider the primary purpose of the paper assessment (OSATS) process was?</td>
<td>Assess Performance</td>
<td>Mainly performance</td>
<td>Both</td>
<td>Mainly feedback</td>
<td>Facilitate Feedback</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tell me your thoughts on the assessment training process you received as part of the peer group? how to use the OSATS/ review video etc? eg. did you need more training, did it help?</td>
<td>Open response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How much specific feedback did you provide?</td>
<td>Occasional comment or general positive comment</td>
<td>Mainly general comments</td>
<td>Both</td>
<td>Mainly detailed suggestions</td>
<td>Detailed suggestions for improvement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How did you deliver the feedback?</td>
<td>Written only</td>
<td>Mainly written</td>
<td>Both</td>
<td>Mainly verbal</td>
<td>Verbal only</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Did you feel you had enough expertise/knowledge to provide feedback?</td>
<td>No</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How useful did you think the feedback you gave to your peer was?</td>
<td>Not useful</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes, very useful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How useful did you think the feedback you gave to your peer was, TO YOUR OWN LEARNING?</td>
<td>Not useful</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes, very useful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>How useful did you think the feedback you received from your peer was?</td>
<td>Not useful</td>
<td>Not really</td>
<td>Neutral</td>
<td>Somewhat</td>
<td>Yes very useful</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any other comments about the peer feedback process?</td>
<td>Open response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do you agree with the implementation as described above, or would you have any different suggestions?</td>
<td>Open response</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
How comfortable would you feel if you required a peer to agree that your performance is good enough before you were reviewed by a faculty member?

<table>
<thead>
<tr>
<th>Not comfortable at all</th>
<th>Not really</th>
<th>Neutral</th>
<th>Somewhat</th>
<th>Yes, very comfortable</th>
</tr>
</thead>
</table>