THE USE OF A SELF-REGULATED LEARNING CONCEPTUAL FRAMEWORK TO INVESTIGATE STUDENTS' ENGAGEMENT WITH INDIVIDUALISED FEEDBACK FROM SUMMATIVE CLINICAL EXAMINATIONS AT A UK MEDICAL SCHOOL

By

BEVERLEY ANN MERRICKS

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF EDUCATION

School of Education
College of Social Sciences
University of Birmingham
June 2017

ABSTRACT

Research regarding students' engagement with feedback from objective structured clinical examinations (OSCEs) is currently limited. Medical students at the University of Birmingham are provided with individual comments from examiners on their performance in summative OSCEs. This thesis explores student engagement with feedback in this context, and if engagement differs depending on the level of performance in the OSCE.

A self-regulated learning (SRL) conceptual framework was devised to illustrate how effective learners incorporate feedback into their learning routines, highlighting the role of reflection in this process. Three empirical studies were conducted to appraise the SRL framework. An interview study (N=11) found that students at different performance levels in the OSCE had contrasting approaches to engaging with their feedback. However, two questionnaire studies (N=180 & 233) with sub-scales to assess the level of student engagement with feedback and their propensity to reflect on their learning, found that the majority of respondents claim to cognitively engage with feedback, whilst a lower number act to improve clinical skills or knowledge. The majority of the respondents were assessed as being reflective learners. Therefore there was no association between performance levels and these attributes. There was a statistically significant association between the engagement and reflection scales, suggesting that these may be related as hypothesised in the SRL conceptual framework, but a causal relationship cannot be asserted from these results.

The findings of this research indicate a gap in the current level of engagement with feedback between cognitive engagement and action. Suggestions are made for activities to ameliorate this situation, which may be relevant for similar contexts.

DEDICATION

I would like to dedicate this thesis to the three most important people in my life. My mother, Pamela Johnson, who has always promoted the benefits of education, and who in her eighth decade continues to encourage her four children to do their best; my husband, Antoine Assaf, who steadfastly believed that I could finish this task and provided support to enable me to have the emotional resilience and the time to complete it; and our daughter, Nadia who found ways to entertain herself and mostly managed to keep out of mischief whilst I was working on this thesis which was started before she was born.

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr Ian Davison who serendipitously inherited me a while ago and Professor Jamie Coleman who later kindly volunteered to jointly supervise my meandering approach to research. They only saw the dismembered bones of this work for some time, but kept the faith until I showed them the whole body. Thank you for your advice and your patience.

I am grateful to the College of Medical and Dental Sciences at the University of Birmingham for sponsoring me to undertake this doctorate and especially thankful to Professor Kate Thomas, Vice Dean and Programme Director MB ChB, for her permission to contact students and for her constant personal encouragement.

I would like to thank the medical students who generously volunteered to share their views by talking to me and completing the questionnaires. Without their support I would not have been able to conduct the empirical research.

Finally, I would like to thank my colleagues, Dr Matt Morgan for being my ‘critical friend’ and Mr Adam Garbett, for providing advice to ensure the questionnaires could be scanned to capture the data effectively and for anonymising data from the Phase 2 questionnaire.

CONTENTS

LIST OF ABBREVIATIONS

BERA	British Education Research Association
BP	Blood pressure
CBD	Case Based Discussion
ECG	Electrocardiogram
EFA	Exploratory Factor Analysis
eOSCE	electronically marked Objective Structured Clinical Examination
EWF	Engagement with Feedback
GEC	Graduate Entry Course
GMC	General Medical Council
GPA	Grade Point Average
HE	Higher Education
HIV	Human immunodeficiency virus
HLE	Higher Level Engagement
KMO	Kaiser-Meyer-Olkin Measure of Sampling Adequacy
LLE	Lower Level Engagement
MB ChB	Combined Medicine and Surgery degree
MCQ	Multiple Choice Questions
NSS	National Student Survey
OMR	Optical mark recognition
OSCA	Objective Structured Clinical Assessment
OSCE	Objective Structured Clinical Examination
PCA	Principal Components Analysis
PM	Personal Mentor
PSA	Prostate-specific antigen
RP	Role Player
RQ	Reflection Questionnaire
SMART	Specific, measurable, attainable, realistic, and time-bound
SRL	Self-regulated learning
TPF	Thinking for Pass or Fail
ZAD	Zone of Actual Development
ZPD	Zone of Proximal Development

CHAPTER 1: INTRODUCTION TO THESIS

1.1 Overview of structure of thesis.

The purpose of this section is to provide an overview of the structure of this thesis and offer a concise summary. However, before this, I think it is useful to include a note about the researcher's 'voice' in this thesis. Readers who are familiar with the physical sciences or medical education literature may be surprised about the use of the first person pronoun throughout this thesis. Although this may not be the usual approach in some other disciples, it is a common approach used in professional Education Doctorates (Burgess et al. 2006, p98) which are situated in social science norms that are explained further in the methodology chapter below.

This first chapter outlines where the idea for this research came from and why it is important to find out if feedback from summative clinical examinations is accessed and used by students. The chapter then goes on to illustrate the specific context of the research and the researcher.

Chapter 2 explores the literature on the theories of learning relevant to higher education, and specifically clinical education. There is a brief discussion on the role of assessment in clinical education. This is followed by a more comprehensive discussion about feedback on learning. The discussion then segues into an exploration of the literature on self-assessment which is an important skill that learners need in order to engage with any feedback. This then leads to consideration of the importance of reflective skills to be able to critically analyse feedback and for a learner to be able to consider if they agree with the feedback giver that there is a deficit in skills or knowledge which they will act on. Finally, the main topic of this thesis is explored; what is known about medical students' propensity to act on feedback provided after summative clinical examinations.

Chapter 3 introduces the self-regulated learning conceptual framework underpinning this research. The conceptual framework aims to describe the cognitive processes or other actions that effective self-regulated learners

undertake and how these relate to engagement with feedback from clinical examinations.

Chapter 4 includes the philosophical approach underpinning this research. The methodological approaches taken in this study are explained and justified. The issues of sampling and ethical considerations are explored. Issues relating to the credibility of the findings from quantitative and qualitative research methods are probed.

Chapter 5 describes the first empirical study; semi-structured interviews with eleven fifth year (Y5) medical students invited via a purposive sampling frame stratified by performance level in the end of fourth year (Y4) examination. These students were invited to discuss their engagement with the range of feedback provided after the Y4 objective structured clinical examination (OSCE) they sat in 2014, but the interview questions mainly focussed on the individualised comments from the examiners. The views and actions of the top performing students were contrasted with those of the bottom performing students in order to examine if the initial outline self-regulated learning conceptual framework was appropriate. The interviews and subsequent analysis of the data were used to inform my understanding of students' current engagement with the feedback. This enabled the development of my conceptual framework for effective self-regulated learning by adding details contextualised to this situation, and informed the next steps in the research.

Chapter 6 describes the process for selecting an existing scale to assess students' level of reflection on their learning during the course. As reflection is described as being an important facet of self-regulated learning, it was hypothesised that effective self-regulated learners would also reflect on their performance in the examination and the feedback provided after it. Kember et al.'s (2000) Reflection Questionnaire was chosen for this purpose.

Chapter 7 describes the second empirical study which was the first phase of the development of a survey instrument to assess student engagement with feedback on a wider scale. The data for this study came from 180 Y4 and Y5 students'

views on their engagement with individual feedback comments from their end of the previous year's (i.e. Y3 and Y4) OSCEs in 2015. Two factors from the Reflection Questionnaire were also included in the administration of a combined survey.

Chapter 8 describes the third empirical study which was the second phase of the development of a survey instrument to assess student engagement with feedback. The data for this study came from 233 Y4 and Y5 students and Y5 graduands' views on their engagement with individual feedback comments from their end of Y3, Y4 or Y5 OSCEs sat in 2016. The same two factors from the Reflection Questionnaire were included in the administration of a combined survey.

Chapter 9 reviews the findings of this thesis and discusses if the initial aims were met and what contribution to knowledge this thesis offers. Some suggestions for further research are given, as well as recommendations specifically for this context but which are generalisable to other contexts. The thesis ends with a brief reflection on my learning journey during the production of the thesis.

1.1.1 Introduction

Archer (2010, p101) believes that 'opportunities to provide feedback must not be missed, including those to impart potentially powerful feedback from high-stakes assessments'. When the Birmingham Medicine and Surgery (MB ChB) undergraduate programme introduced an OSCE at the end of the fourth year of study in 2012-13, it followed the existing third and fifth year processes for provision of generic and comparative feedback to students but improved on this by also providing personalised feedback comments to students from their examiners on their performance at each of twelve OSCE stations (questions). This was despite the logistical difficulties presented by a large cohort of approximately 400 students. Following this the provision of individual feedback comments on performance in the clinical examinations was implemented for Y5 in 2014 and Y3 in 2015.

Medical students do not necessarily engage with feedback (Sinclair and Cleland, 2007, Harrison et al., 2014). It has been shown that the top performing students are more likely to use feedback effectively than poorly performing students, who

arguably would benefit more from doing so (Harrison et al., 2013). This incongruous behaviour may be explained by self-regulated learning theory (Zimmerman, 1989). Higher performing students may be more effective self-regulated learners who are more likely to integrate feedback into their usual learning routines.

It was not known if, or how, medical students in the context of this study engaged with their personalised feedback in order to confirm or improve specific knowledge and skills. It is important that medical students learn to take cognisance of all types of feedback on their clinical performance because this is a life-long skill for independent clinical practitioners and, ultimately, is in the interests of patient safety (Health Education England, 2016). The General Medical Council (GMC), which is the statutory regulatory body responsible for oversight of the training and conduct of doctors at all stages of their professional development, has explicitly stated that they expect students to be encouraged to act on feedback:

R3.13 Learners must receive regular, constructive and meaningful feedback on their performance, development and progress at appropriate points in their medical course or training programme, and be encouraged to act on it. Feedback should come from educators, other doctors, health and social care professionals and, where possible, patients, families and carers (GMC Promoting Excellence 2015, p26).

In line with many other higher education institutions, the impetus for the provision of additional feedback for students on the MB ChB programme in Birmingham was driven by a desire to improve the National Student Survey (NSS) scores on assessment and feedback as well as to improve the educational experience of students. Holmes (2015, p3) notes '…it is within every university's best interests to achieve the highest NSS scores as possible'. The potential effect of the improvement to feedback on the NSS scores from 2013 to 2016 is discussed in Chapter 9.

1.1.2 Research topic, objectives and questions

At the time of the introduction of the new Y4 OSCE in 2013, I was satisfied that the logistical and technical difficulties of providing individual comments had been

overcome and I had briefed the examiners on what was expected of them to provide the feedback. I did not consider if students would access and engage with the personalised feedback. I viewed this from my own perspective which presumed that most medical students would be motivated to at least read these personalised comments and potentially use them to improve on any weaknesses in clinical knowledge and skills.

I serendipitously read Harrison et al.'s (2013) report on Keele University medical students' engagement with web based feedback from a summative assessment soon after it was published. This described how the top-performing students accessed the most pages of information and revisited the pages, whilst the just passing students engaged with this information the least. The challenge presented was to implement strategies to encourage the more poorly-performing students to engage with feedback with the aim of improving their performance. Harrison et al. concluded that for reasons of feasibility, they had not provided individual examiner comments to students, and that if they had, the outcome may have been different (p742).

Reading this paper sparked a 'light-bulb' moment which led me to reflect on the situation at my institution. I wanted to know if there was a similar pattern of differential engagement with feedback, depending on level of performance, for medical students at Birmingham. I initially believed that if this was the case we, as a faculty, should devise a teaching intervention or other strategy to remedy this, so that students at lower levels of performance could be encouraged to engage with the feedback and improve their knowledge and skills. This concern about potential lack of engagement was not only motivated by a desire to meet the GMC's regulatory requirements, but to help weaker students develop their ability to utilise feedback in the same way that time is apportioned in the curriculum for the development of other skills which have an internal locus of control, like professional behaviour and attitudes.

A review of the relevant literature at that time suggested that there were very few reports about student engagement with individualised feedback after summative clinical examinations. This thesis seeks to answer questions about whether

students engaged with this feedback, and whether the performance level of medical students in their end of year summative clinical examinations is associated with their level of engagement with the individual feedback comments provided and/or their propensity to reflect on learning events generally. The study sought to view and answer these questions through the lens of a self-regulated learning conceptual framework which could have wider applicability, rather than an evaluation study of the quality of local provision.

1.2 Research Context

The section below gives background information about the MB ChB degree course and the summative clinical examinations that give rise to the feedback discussed in this thesis. It then goes on to discuss the context of how the individual feedback comments are produced. Next exemplars of the feedback from Y4 are provided. Finally the contextual and epistemological position of the researcher is described in order to make overt the factors influencing key decisions made about why and how this research was conducted.

1.2.1 Batchelor of Medicine and Surgery (MB ChB) course at Birmingham

The MB ChB is a professional undergraduate course, which is externally validated by a statutory body, the General Medical Council. Assessments on this course are complex to organise given that there are about 400 students per year on a five year course and 40 graduate entry students per year on a four year course. Students are taught and assessed by small numbers of biological and social scientists employed by the University of Birmingham and large numbers of NHS-employed clinicians, practicing in numerous clinical sites and settings across the West Midlands.

1.2.2 Objective Structured Clinical Examinations (OSCEs)

The Objective Structured Clinical Examination (OSCE) was initially devised by Harden and Gleeson in Scotland in the 1970's (Harden and Gleeson, 1979). It is now a commonly used format for assessing medical students' clinical and communication skills at undergraduate and similarly for doctors at postgraduate level in the UK and in the rest of the world (Patricio et al. 2013). This format of

assessment is also used by other healthcare professionals such as nurses (Rushforth, 2007), and dentists (Mossey et al. 2001).

Khan et al. (2013a, p1440) proposed a definition of the OSCE as

> 'An assessment tool based on the principles of objectivity and standardisation, in which the candidates move through a series of time-limited stations in a circuit for the purposes of assessment of professional performance in a simulated environment. At each station candidates are assessed and marked against standardised scoring rubrics by trained assessors'.

The OSCE is objective in that it aims to minimise variation caused by the examiner or patient or simulated patient (role player) and attempts to only examine the variation in the performance of the candidates. It is structured in that there is a timed, well-defined interaction between candidate and examiner and an associated clear marking scheme. The questions set aim to test specific clinical skills including communication, patient body systems examination, clinical procedural skills, diagnosis and management of clinical presentations.

An OSCE typically comprises a circuit of timed physical 'stations' that each candidate visits in turn and is asked to demonstrate a range of specific knowledge, skills or behaviour. Usually the same number of students is assessed as the number of stations or questions in an OSCE circuit and it does not matter which station a candidate starts at, during the examination they rotate through all the stations or questions.

The OSCE can be a reliable, valid and defensible method of assessing a range of clinical competencies if it is well designed and implemented (Boursicot et al. 2010). However, as with any other method of assessment, it has limitations including the following. There is a risk of compartmentalisation of separate clinical skills rather than the demonstration of an authentic integrated approach (Nestel et al. 2011). The OSCE is an assessment of performance in a simulated setting at the 'shows how' level rather than the 'does' level of for example, work-place based assessments, which are at the top of Miller's pyramid of assessment of clinical

skills (Miller, 1990). Examiners may suffer fatigue from doing or saying exactly the same thing multiple times and this may lead to a decrease in objectivity and standardisation (Humphris and Kaney, 2001). Developing high quality OSCE questions takes time and effort (Khan et al. 2013b). It is a resource intensive, expensive method of assessment (Reznick et al. 1993).

1.2.3 Individual feedback comments from the OSCEs at Birmingham

In the context of this study, the examiners do not know the candidates and make a snapshot judgement of each student's performance based on the expectations for a minimally competent end of Y3 or Y4 student, or for Y5, a junior doctor on their first day of practice. These expectations for competency are explicitly provided for examiners in 'anchor statements' which are word pictures for each of four performance levels (very poor/not done; unsatisfactory; satisfactory; very good) provided for that question. Examiners do not know the numerical marks they are awarding as these are not shown on the mark sheet. Instead, examiners indicate one of the four performance categories for each task or question on a mark sheet which can be read by Remark 8 Optical Character Recognition software using a Fujitsu FI 5530C scanner. These tasks or questions may have differently weighted marks which have been allocated by the question author who is a specialist in that clinical topic. Examiners are also instructed to provide one or two polite, specific, actionable and legible handwritten feedback comments under each of the headings 'What did you do well?' and 'What could you improve?' An example question and marksheet is included as Appendix 1.

Each completed mark sheet is scanned to extract the data and the individual feedback comments. The marks for each student are calculated per question and calibrated against the question pass mark which is set using the Borderline Group Method (Livingston and Zieky, 1982). The overall mark awarded per question is calibrated to the standard 50% pass mark before being released to the students. Students must pass a minimum number of stations (e.g. for Y4 this is 8 out of 12) as well as achieving at least 50% for the examination overall. The scanned individual feedback comments are made available to each student as image files via a web-based virtual learning environment. The examiner feedback comments may be briefly reviewed by the examination site organisers, or by the scanning

machine operator, but no robust quality assurance checks are made before the comments are released to students.

There is a delay of about four weeks in releasing feedback after the OSCE to enable data processing and checking of examination results followed by ratification by the Examination Board. Marks are normally released shortly before the other types of feedback, so students know if they have passed or not, and this may influence whether they look at any of the other feedback.

1.2.4 Examples of all the types of feedback provided after OSCEs

As an illustration of the feedback provided, information about the end of Y4 OSCE at Birmingham is shown below. The Y4 OSCE comprises a carousel of twelve stations (questions), each seven minutes long with a three minute gap so that students can read the question instructions outside their next station whilst the examiners complete a mark sheet for their previous candidate which includes writing the feedback comments.

Five types of feedback are provided to students after the Y4 OSCE:

Individual information

1. The calibrated score for each of the twelve questions the student sat and their overall average score.
2. Handwritten comments provided by each of the twelve examiners under the headings ‘What did you do well?’ and ‘What could you improve?’

Contextual/generic information

3. The mean calibrated score for each of the 36 questions used during the two day examination period.
4. A histogram of the cohort’s overall calibrated OSCE scores
5. Generic comments synthesised from comments collected from all examiners from different hospital examination sites for each question.

Students who fail the main sit of the OSCE also receive additional feedback and advice for improvement during an individual interview with the Year Lead.

Table 1.1 Examples of individual feedback comments for roleplayed question on Primary HIV 2014

(NOTE: Transcribed exactly from the handwritten comments of different examiners.)

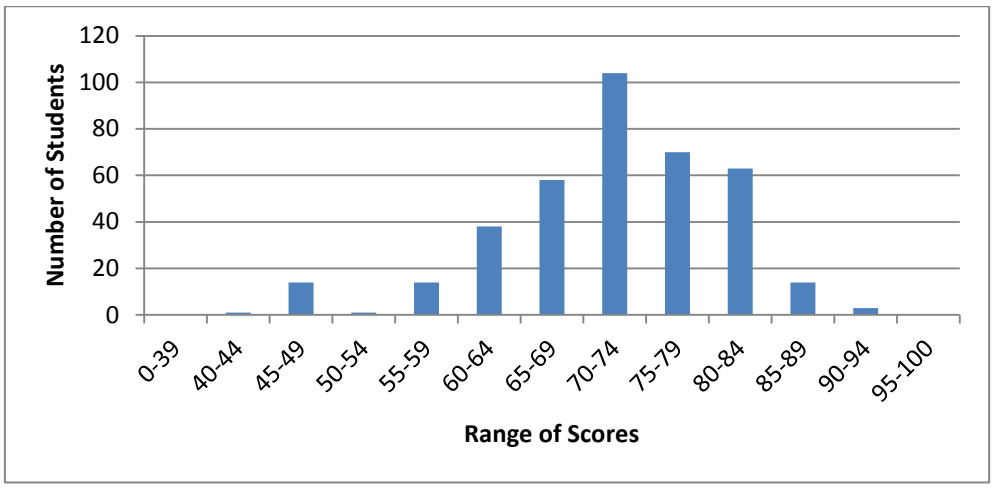
What did you do well?	What could you improve?
Empathy, knowledge	Do not use jargon. Concentrate on detailed sexual history and HIV risk assessment. Both were modest.
Reassured patient of confidentiality and there was excellent use of language and choice of words that was inherently reassuring and understanding. A very efficient consultation	Continue to maintain this high standard of listening and professionalism
Strong questioning and body language. Good eye contact with patient. Very logical sequence of asking questions. Very considerate and taking care of confidentiality	Needs to be more confident in presentation and answering differentials
Very amiable, approachable	Was repetitive many times Kept apologising couple of times Needs to work on more focussed history taking skills.
Good communication skills generally. Obtained most of the information.	Can improve on active listening (repeated some questions, answers to which were already known). Can improve on fluency Used word 'worried' over 5 times. Need to control the nerves Very occasional use of jargon
Explored well sexual history Good interaction with patient	Structure better opening questions to patients from general to personal.
Knowledge and empathy excellent	Try and feel less awkward when asking sensitive questions.
Sexual history was organised, although some details missing	Jargon 'ask about a blood borne infection risk assessment'. Be slightly more empathetic & organised
Good history	Explain difference between HIV +ve & seroconversion in diagnosis
Did not use jargon.	Pick up on pt cues. She said twice she was worried and you did not explore this.

Table 1.2: Mean calibrated score for each of the 36 questions used in the 2014 OSCE

	Tuesday AM	Tuesday PM	Wednesday AM	Wednesday PM
Station 1: RP 1	HIV 1 73%	Psoriasis 64%	Psoriasis 64%	HIV 1 73%
Station 2: RP 2	Sick Sinus Syndrome 77%	Facial Pain 75%	Facial Pain 75%	Sick Sinus Syndrome 77%
Station 3: RP 3	TIA 75%	GCA 75%	GCA 75%	PSA screening 70%
Station 4: RP 4	Altered Bowel Habit 79%	Altered Bowel Habit 79%	PSA screening 70%	TIA 75%
Station 5: CBD 1	Bone Metastases 72%	HIV 2 72%	HIV 2 72%	Dyspepsia 79%
Station 6: CBD 2	Epistaxis 82%	Diabetes 84%	Delirium 75%	Uveitis 82%
Station 7: CBD 3	Melanoma 81%	Bone Metastases 72%	Dyspepsia 79%	Melanoma 81%
Station 8: CBD 4	Uveitis 82%	Epistaxis 82%	Diabetes 84%	Delirium 75%
Station 9: PP 1	Airway Maintenance 80%	BP and ECG 84%	BP and ECG 84%	Airway Maintenance 80%
Station 10: PP 2	Shoulder Exam 79%	Knee Exam 76%	Shoulder Exam 79%	Knee Exam 76%
Station 11: PP 3	Urine Dipstick 85%	Urine Dipstick 85%	Ear Examination 70%	Ear Examination 70%
Station 12: PP 4	Peak Flow 82%	Oxygen Therapy 82%	Oxygen Therapy 82%	Peak Flow 82%

NOTE: RP = Role played station, CBD = Case based discussion station, PP = Practical Procedure station

Figure 1.1: Histogram of the cohort's overall calibrated OSCE scores 2014



Note: Students who achieve ≥50% overall but pass less than 8 stations have their overall marks 'capped' at 49% to show that they failed at this attempt

Table 1.3: Example of generic comments for role played question 'Possible HIV exposure'

In general what did students do well on this station?	In general what did students do poorly or not do on this station?
• The general standard of communication was good and students gave good explanations laying out context and why sensitive and personal questions needed to be asked. • Most got the primary diagnosis of HIV and knew about testing.	• Too much use of medical jargon generally. • More detail needed in sexual history and particularly about partner - many forgot to check their gender. • Many struggled with differential diagnoses such as syphilis. HIV seroconversion was rarely discussed.

1.2.5 Positionality of the researcher

As well as providing information about the context in which the research was conducted, it is pertinent to include some biographical and contextual information about the researcher. My job title is Education Development Specialist and I am currently responsible for the delivery and development of assessments on the MB CHB programme at the University of Birmingham. I am a middle aged woman who has worked in medical education at Birmingham for 19 years, and was employed by another University in a central educational development role for 5 years prior to this. Although I share a 'community of practice' (Lave and Wenger, 1991) with the students and their teachers, I am not a health care professional. I am a personal tutor for a small group of medical students and give plenary lectures on assessments, but I do not usually have contact with students, so I do not have an in depth understanding of the students' day to day lived experiences.

Being associated with the subjects of this research has the related benefits of relatively easy access to potential respondents and knowledge of the learning and assessment processes they are subject to, as well as understanding their jargon. However, this also brings the potential pitfall of my studying student engagement with feedback with preconceived views or expectations which, if not consciously guarded against, may lead to bias due to my personal involvement in all aspects of the clinical assessments and the provision of the feedback comments.

1.3 Summary

This chapter introduces the topic of feedback after summative clinical assessments and alludes to the fact that recipients of this feedback may not necessarily engage with it. The chapter outlines the research questions this study aims to provide evidence about, and provides a guide to the topics that will be included in greater detail later in this thesis. It provides contextual information so that the reader can understand how the clinical examination feedback is generated and gives examples of what it is. It introduces the researcher as a person in the real world and the notion that the researcher's positionality and approach to research influences everything that is presented in this thesis.

CHAPTER 2: FOCUSSED LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to provide the reader with a thematic survey of the relevant literature to give sufficient background information to put into context the topic of students' engagement with written feedback after clinical examinations. I will begin the chapter by outlining the broad theories of learning which are most pertinent to this study. However, the theory of learning which specifically underpins this thesis, self-regulated learning and the related topic of reflection on learning will be covered in Chapter 3. Next, I will briefly consider the effect assessment has on learning. I will then discuss themes within feedback on learning which are relevant for this study. I will then introduce the topic of self-assessment of performance, which is an essential component of engaging with feedback. Following this, I will describe my literature searching technique for relevant published work on students' engagement with feedback from clinical examinations and examine what is currently known about this topic. Finally I will summarise this focussed literature review chapter.

Before I review the literature as described above, there follows a brief mention of topics which are related to this field of study but will not be included here, for reasons of delineation and brevity. Firstly, issues relating to the quality of the written feedback comments provided and how the quality may be assessed will not be included (White and Sharma 2012; Hughes et al. 2015; Bartlett et al. 2017). This is a very pertinent area with the potential to influence engagement, but this thesis will mainly explore engagement with feedback independent of the quality of the feedback provided. Secondly, although the role of peer assessment and peer feedback in student learning is an important and growing area of literature, it will not be explored in this thesis. It has been noted that the giving and receiving of feedback from peers is of great help to learners as this helps to make tangible the expected standards and provide bench marks by showing what standard is possible for other students to achieve, and inform self-assessment (Falchikov 2007; Ladyshewsky 2013; Nicol et al. 2014).

2.2 Theories of learning and their relationship to feedback

Swanwick and Buckley note that the ultimate aim of medical education is to provide a knowledgeable, skilled and up-to-date professional workforce who put patient care above self-interest and who maintain and develop their range of expertise over the course of their career (2010, pxv). There are a number of theories to conceptualise how people learn, not only in medical education, but at different ages, levels of academic provision and learning environments. Three relevant, overarching ones; behaviourism, cognitivism and constructivism will be considered in brief here, along with three related theories that have developed from these; social-cognitivism, social-constructivism and socio-culturalism.

Behaviourism

Behaviourism is one of the earlier theories of learning (Thorndike 1911; Watson 1913; Pavlov 1927; Skinner 1958). Behaviourism conceptualises learning as the acquisition of desired behaviour in response to external stimuli from the environment. Teaching is aimed at shaping the responses of learners through modelling, demonstration and reinforcement of the targeted response via direct instruction from a teacher who maintains control of the pace, sequence and content of the lesson (Palincsar 1998, p346). Correct demonstration of the desired behaviour leads to immediate feedback in the form of positive or negative reinforcement. Mann believes that the importance accorded to the provision of feedback on learning originated in behaviourism (2011, p45).

Behaviourism as an explanation of how and why learners progress has been largely superseded by other theories, mainly because behaviourism does not take into account the learner's cognitive processes or the influence of social interactions within the learning environment. However, it still has a role in medical education in learning skills where there is a defined correct approach to doing things in order to ensure patient safety, (for example aseptic techniques) or skills which are undertaken under stressful circumstances so the learned behaviour becomes automatic, (for example when attempting to resuscitate a patient). The approach to feedback in a behaviourist learning paradigm is corrective, i.e. pointing out what was done correctly or wrongly and should be immediate. This

approach to feedback may be appropriate for skills training (Ker and Bradley 2010, p169).

Cognitivism

In the middle of the twentieth century there was a move away from behaviourism to learning theories and models influenced by psychology which emphasised the importance of complex cognitive processes such as thinking, problem solving, language, concept formation and information processing instead of overt, observable behaviour (Ertmer and Newby,1993). Cognitive theories focus on students' learning processes in terms of how knowledge or information is received, organised, stored as memory, and retrieved when required. Knowledge is seen as schema or symbolic mental constructions. Learning is defined as change in a learner's schemata (Kirschner et al., 2006). In this conceptualisation, new knowledge builds upon prior knowledge to develop expertise. Learners need to actively participate in order to make individual meaning from their learning experiences. Changes in behaviour may be observed, but only as an indication of what is taking place in the learner's mind. In cognitive models of learning, the feedback process starts with giving the feedback from the teacher which is then processed by the learner (Thurlings et al. 2013). The role of feedback is to guide and support accurate mental connections or schemata (Kirschner et al., 2006).

Social-Cognitivism

Facets of behavioural and cognitive approaches to learning were united in social-cognitive theory as propounded by Bandura (1977). This theory acknowledges the importance of the social and interactive aspects of learning (Kaufman and Mann 2010, p18). Social cognitive theory views the learner as an active agent in learning which is influenced by that individual's goals, attitudes, values, knowledge and experience. Learners are able to monitor aspects of their own progress towards their learning goals (Mann 2011, p 63). As well as acknowledging the role of learner self-efficacy or self-regulation, social-cognitive theory also highlights the importance of the metacognitive skill of learner self-reflection to gain understanding of themselves, their behaviour and the environment (Kaufman and Mann 2010, p20). Self-regulated learning as a model to conceptualise what effective learners do, was developed from Bandura's earlier work on social-

cognitive theory (Zimmerman 2013, p135). Self-regulated learning will be further discussed in Chapter 3.

Feedback processes start within the learner who has set learning goals and can start to self-evaluate to compare their planned and actual achievement. The provision of external feedback on performance can increase the learner's motivation to enhance their performance effort (Bandura and Cervone, 1983).

Constructivism

Constructivism is an approach to conceptualising learning which gained credence in the 1970's and 1980's. It is similar to cognitivism in that both conceive of learning as a mental activity, but constructivism equates learning with creating meaning from experience. Constructivism is based on the principle that learners are not passive recipients of information, but that they actively construct their knowledge by interaction with their environment and reorganisation of their mental schema. Learners are viewed as sense-makers, who interpret new information. While there are different views on constructivism, what they have in common is a learner-centred approach whereby the teacher becomes a cognitive guide for learning and not a knowledge transmitter (Palincsar 1998, p348).

Effective feedback for this approach to learning should be directed at the task, be specific, detailed and clear. It should help learners to close the gap between their actual and the desired outcomes (Thurlings et al. 2013, p9). Feedback should enable the learner to analyse their actions and their understanding and plan for future learning (Dennick, 2012, p622).

Social-constructivism

Later in the twentieth century, the constructivist view of learning was further developed to emphasise the influence of the social and cultural milieu on what is learned and the process of how it is learned via social interaction. This compares with the constructivist view of learning as being a mental process, with knowledge being considered as self-sufficient and independent of the contexts in which it is learned or applied. The pioneering work of Vygotsky in Russia from the 1920's

and early 1930's (he died in 1934) was translated to English (Vygotsky, 1978) and his views on the social nature of learning were more widely disseminated. In social-constructivism, cognition and learning are understood as interactions between the individual and a situation; knowledge is considered as situated and is a product of the activity, context and culture in which it is formed and used and is a shared, rather than an individual, endeavour. Vygotsky drew attention to the concepts and tools a 'more knowledgeable other' (e.g. parent, tutor, peer) use to mediate the learning of another person and stressed the importance of language (shared talk) in the process. He introduced the concept of the 'zone of proximal development' (ZPD) which refers to what a learner can do with the assistance or guidance of a more knowledgeable other. This ZPD is held to be where learning actually takes place. This is contrasted with the learner's 'zone of actual development' (ZAD), which is what the learner can do alone (Morris and Blaney 2010, p74).

Viewed through the lens of the ZPD, feedback from a more knowledgeable other should use appropriate language, or artefacts, to scaffold the learner through the ZPD and into a ZAD, and so feedback needs to be at the correct level for that learner's development so it can be internalised and accepted as being meaningful.

Socio-Culturalism

Socio-cultural perspectives on learning also conceptualise learning as an activity that is essentially communal and located in a society, rather than only in the minds of individuals. Socio-cultural learning theories such as 'communities of practice' acknowledge the importance of interactions for both individual and collective learning in workplaces. This is also known as 'situated learning'. There are areas of overlap with Vygotsky's work but the tenets of communities of practice were developed by Lave and Wenger (1991) from an ethnographic rather than educational or psychological paradigm. Lave and Wenger see learning, meaning and professional identity as inextricably tied up with one another and with work-related practice from that specific context (Yardley et al. 2012, p e107).

Communities of practice describe the activities of a group of people who come together in pursuit of a shared enterprise. In this conceptualisation of learning, new learners begin at the periphery of the community of practice, observing and performing basic tasks. These learners are known as 'legitimate, peripheral participants' as they have a mandate to be actively involved as part of the community. Through continued participation, engagement and by assuming increasing responsibility, individual learners acquire the roles, skills, norms and values of the community, and in turn, transform and continue the community (Mann 2011, p64). As learning occurs through social interaction, learners use the language and jargon of the community. Although the conception of learning medicine via communities of practice is appealing in certain regards, it has been criticised for having weaknesses. For example in addressing individual variations in accessing learning in a work-place and the ways in which the 'old timers' continue to learn in the work-place (Morris and Blaney 2010, p76).

Feedback to learners in a community of practice should be encouraged from all members (who will not only be medics) and be on all aspects of the norms of the community (not just knowledge and skills, but professional behaviours, dress, use of language), e.g. multi-source feedback. Feedback should be perceived as credible and well intentioned. It should be frequently provided, accurate, contain concrete and specific information, and be directed at the task (Thurlings et al. 2013, p10).

Theories of learning employed in this context

Application of a learning theory is mediated by the course of study or curriculum and the educational provider's associated rules, structures and processes under which the learning theory is operationalised. The University of Birmingham supports an 'enquiry-based approach to learning' which entails 'giving its students the support they need to become independent and self-motivated learners', rather than advocating a specific theory of learning (University of Birmingham, 2017). It can be argued that students in the context of this study may experience elements of all the learning theories described above at different times over the five years of the course and in different learning situations (e.g. full cohort lectures, small group tutorials, self-directed learning using books, electronic media or other artefacts,

alone or with peers and in various clinical environments with clinical tutors and patients). Medical students need to learn a diverse range of knowledge, skills and professional behaviours and are guided by a large cadre of tutors, whose primary duties are to provide healthcare and many of whom are not trained as teachers. Similarly, students also need to learn to negotiate and navigate a range of types and styles of feedback.

2.3 Assessment effects on learning

Boud and Falchikov contend that 'Assessment, rather than teaching, has a major influence on students' learning. It directs attention to what is important. It acts as an incentive for study' (2007, p4). A number of authors have provided frameworks for assessment conditions which support learning in order to maximise the learning potential of assessment (e.g. Barr and Tagg 1995; Rust 2002; Gibbs and Simpson 2004; Rust et al. 2005; Norcini et al. 2011). These are mainly framed in social-constructivist learning perspectives and include elements such as:

- Ensuring the whole course is 'constructively aligned' so that the learning outcomes, the learning and teaching methods and the assessment methods are explicitly interrelated (Biggs, 1996)
- The provision of clearly defined and explicit assessment processes and criteria are provided to learners and tutors.
- There are opportunities for active engagement with the assessment criteria and standards.
- Feedback is an integral component of assessment.
  (adapted from Rust et al. 2005).

Examples have been provided in the literature which discuss how changes to specific assessment arrangements appear to have directed students to more effective learning strategies, for example Newble and Jaeger (1983); Rudland et al. (2008); Norman et al. (2010). Research in cognitive psychology has shown that assessments can directly affect learning by promoting better retention of information, a phenomenon known as the testing effect (Larsen et al. 2008).

Assessment formats may be categorised as summative or formative. Formative assessment is concerned with how judgements about the quality of a student's responses (performances or work) can be used to shape and improve the learner's competence by short-circuiting the randomness and inefficacy of trial and error learning. Summative learning is concerned with summing up the achievement status of a student and is geared up towards reporting at the end of a course of study, especially for purposes of progression on the course, ranking or certification (Sadler 1989, p120).

However, the distinction between these two are now blurred; as highlighted by Taras 'all assessment begins with summative assessment (which is a judgement) and that formative assessment is in fact summative assessment plus feedback which is used by the learner' (2005, p466). Many courses in higher education now provide feedback after summative assessment (e.g. Harrison et al. 2013; Gierl et al. 2014; Carruthers, 2015).

A more contemporary classification is to refer to the *intended effects* of an assessment as being 'of', 'for' or 'as' learning. Assessment of learning is synonymous with summative assessment. It is the use of an activity to measure, record, report on a student's level of achievement judged against specific learning expectations or against other test takers' results. There are consequences for the results and there may be external scrutiny with regards to public safety or regulated standards (Wiliam, 2011). Assessment for learning is 'the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there' and is 'part of everyday practice by students, teachers and peers that seeks, reflects upon and responds to information from dialogue, demonstration and observation in ways that enhance ongoing learning' (Broadfoot et al., 2002). Assessment as learning 'incorporates the programmatic and curricular implications of the assessment process that might involve looking back from the assessment system to the curricular experiences and the goals of the educational program. Assessment as learning might also consider incorporating the effects of relationships between assessors and students' (Sklar, 2017). There is growing interest in the implementation of a programmatic approach to assessment in

medical education which entails elements of all three of these approaches (van der Vleuten et al. 2005; van der Vleuten et al. 2012; Bok et al. 2013; Heeneman et al. 2015; Eva et al. 2016).

Although Watling (2016) refers to the 'uneasy alliance of assessment and feedback' there is increasing evidence about the necessity for learning and assessment activities to be carefully and explicitly aligned and interwoven and the need for all agents involved in the learning process (tutors, students, peers) to be enabled to fully understand their roles, especially the role of the student in actively engaging with feedback on their learning activities.

2.4 Feedback on learning

There is a massive corpus of historical and contemporary published work covering every aspect of feedback on learning, from both the provider's and receiver's perspectives. In order to define the boundary about what should be included in this brief review, I considered if the evidence from medical education literature would be more relevant to the learners in this context than the literature framed in a wider higher education perspective. Much of the research on feedback in the medical education literature focuses on feedback provided on activities in the clinical workplace rather than after summative clinical examinations, and has postgraduate doctors in training as the subjects, rather than medical students. As discussed above, assessment and feedback are regarded as overlapping educational activities within social-constructivist learning frameworks and much of the research into social constructivist assessment and feedback practices is located within or across other higher education disciplines.

I decided to focus on the key aspects discussed below that could potentially affect engagement with feedback after clinical examinations, and include evidence from both medical education and wider higher education, as appropriate. For readers interested in other aspects of feedback, Evans (2013) provides a useful systematic review of the work on feedback after assessment, including a thematic overview of the research into principles of effective assessment practice.

Definitions of feedback on learning

Boud and Molloy (2013) provide a contemporary definition for feedback on learning in general which is explicitly located in a social-constructivist learning framework;

> 'Feedback is a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standards for any given work and the qualities of the work itself, in order to generate improved work' (2013, p6)

Feedback in clinical education has been defined in more behaviouristic terms as 'specific information about the comparison between a trainee's observed performance and a standard, given with the intent to improve the trainee's performance' (Van der Ridder, 2008 p193)

Has feedback been shown to have an effect on learning?

O'Donovan et al. state 'When considering the assessment cycle… the literature is clear that potentially the most powerful and potent part of that system, when it comes to improving the future learning of the student, is the feedback the student receives' (2016, p1). Several authors have conducted systematic and meta-analyses of studies on 'feedback interventions' seeking to judge the effect size on learning gains. Kluger and DeNisi (1996) conducted an analysis of 607 studies and report an overall positive effect. However, they caution against reporting average effect sizes from meta-analyses as there was a wide range of effect sizes, including about a third of the studies which reported a negative effect on learning gain. Hattie and Timperley (2007) summarise previous systematic reviews undertaken by Hattie of research in mainly school level education and concluded that feedback is the most powerful single influence that makes a difference to student achievement. Black and Wiliam (1998) also conducted a systematic review and reported that feedback can have very large and consistently positive effects on learning, compared to other aspects of interventions designed to improve learning. In the context of medical education, Veloski et al. (2006) conducted a systematic review of 41 studies which evaluated the independent effect of feedback on physician performance. They found 32 studies which demonstrated a positive impact.

Van de Ridder et al. (2015) conducted a meta-review of 46 reviews and meta-analyses in order to understand the different variables that could affect the processes and outcomes of feedback in settings relevant to medical education. They identified 33 variables which they grouped as task standards and task performance; task observation and interpretation; feedback communication; feedback reception and interpretation; and all of which could influence the feedback process, feedback effect or both (2015, p662).

What is effective feedback?

Hattie and Timperley (2007) propose a model of effective feedback which specifies three kinds of questions that the feedback must answer: Where am I going (understanding of learning outcomes or goals – feed-up); How am I going (progress made towards the learning outcomes or goals – feedback); Where to next (what activities need to be done to make progress – feedforward) (2007, p88). Each of these three questions operates at four levels; feedback about the task; feedback about the processing of the task; feedback about self-regulation; and feedback about the self as a person. They demonstrate that feedback about the self is the least effective form of feedback and that feedback aimed at self-regulation or the processing of the task assist with deep mental processing (2007, p90). Hattie and Timperley also discuss the effects of the timing of the feedback with regard to each of the four levels and the effects of positive (things that were done well) and negative (things that require improvement) feedback at each of these levels.

A number of guidelines for giving feedback specific to clinical education have been published (Ende 1983; Hewson and Little 1998; Henderson et al. 2005; Cantillon and Sargeant 2008; Ramani and Krackov 2012; Lefroy et al. 2015). There is also a commensurate set of guidelines aimed at informing clinical students how to make best use of feedback (van der Leeuw and Slootweg, 2013).

A commonly cited framework from higher education which is overtly located in a social-constructivist framework and aims to enhance self-regulated learning is

Nicol and Macfarlane-Dick's (2006, p205) 'Seven principles of good feedback practice'. They state that good feedback:

(1) Helps clarify what good performance is (in terms of goals, criteria or standards).
(2) Facilitates the development of self-assessment and reflection in learning.
(3) Delivers high-quality information to students about their learning.
(4) Encourages teacher and peer dialogue around learning.
(5) Encourages positive motivational beliefs and self-esteem.
(6) Provides opportunities to close the gap between current and desired performance.
(7) Provides information to teachers that can be used to help shape teaching.

What is effective engagement with feedback?

Sadler posits that in order to be able to engage effectively with feedback, students need to be trained in how to interpret feedback, how to make connections between the feedback and the characteristics of their own work and how they can improve their future work (1989, p79). Price et al. (2010) take up this theme, stating that the learner is in the best position to judge the effectiveness of the feedback, but may not always recognise the benefits it provides. The 'pedagogic literacy of students is key to the evaluation of feedback and related processes' (2010, p277). This view is shared by Bowen et al. (2017) who feel the way to improve student engagement with feedback is to improve students' 'feedback literacy' skills.

Much of the literature on feedback explores student satisfaction with feedback after assessment, rather than changes in students' learning behaviour (Price et al. 2011). However, Winstone et al. (2017) provide a thematic review of 195 papers relating to feedback in higher education settings and offer the SAGE taxonomy to describe the elements of what they term 'proactive recipience of feedback' to denote the active engagement of the learner with the feedback. This recipience framework entails self-appraisal; assessment literacy; goal-setting, self-regulation; engagement and motivation. This was published after I had conducted my empirical research and completed the self-regulated learning conceptual framework for effective engagement with feedback presented in Chapters 3 and 5.

Are there possible barriers to engagement with feedback in my context?
Barriers to engagement with feedback discussed in the literature include the lack of perceived credibility of the feedback provider by the receiver in terms of the provider's characteristics such as experience or knowledge level, or behaviour such as whether they directly observed the performance they were assessing and their style of communication (Bing-You et al. 1997; Veloski et al. 2006; Murdoch-Eaton and Sargeant 2012; Watling et al. 2012; van der Ridder et al. 2015; Telio et al. 2016).

The approach to providing feedback to students in this study is via what Carless (2015) conceptualised as the 'old paradigm' of feedback as monologic information transfer, which is in contrast to the 'new paradigm' of feedback as dialogic interaction. There may be a need for dialogue between the feedback provider and receiver so that the learner can explore and clarify the meaning to make sense of the feedback in their own terms (Nicol 2010; Beaumont et al. 2011; Orsmond et al. 2013; Yang and Carless 2013; Blair et al. 2014; Ajjawi and Boud 2017). This attention to the role of dialogue in a feedback interaction has been extended to calls for a reconceptualization or re-positioning of the learner's role in feedback from passive recipient to an active seeker of feedback and meaning maker within a dialogic framework in line with constructivist theories of learning. This is summed up by Ajjawi; 'Viewing feedback as dialogic, situated and relational or social, rather than individualistic and transmitted, as well as recognising its wider role in promoting self-regulation and professional socialisation would be more helpful' (2012, p1018).

In the context of this study, the students already know if they have passed the OSCE before they are advised that the feedback is available, so they may not be motivated to access the feedback if they have passed (Black and Wiliam 1998; Rust 2002; Irwin et al. 2013). If students access the feedback, they may not understand the comments (Orsmond and Merry 2011; Winstone et al. 2016). Additionally, they may not appreciate its potential role as feedforward for improving future practice (Murdoch-Eaton and Sargeant 2012; Harrison et al. 2014; Bowen et al. 2017). There may be differential engagement depending on the performance level of the student (Sinclair and Cleland 2007; Harrison et al. 2013). Various

authors have presented findings which highlight the role of negative emotions in influencing learners' engagement with feedback (Sargeant et al. 2008; Bing-You and Trowbridge 2009; Eva et al. 2012; Molloy et al. 2013; Urquhart et al. 2014; Van der Leeuw 2014; Bynum 2015). The framing or valence of the feedback (positive or negative) or other perceptual variables, such as attributions of success or failure may also have an impact on the student's propensity to engage with their feedback comments (Fishbach et al. 2010; Kluger and van Dijik 2010; Watling et al. 2013; Crommelinck and Anseel 2013; van der Ridder et al. 2015)

Are there factors which potentially promote engagement with feedback in my context?

The fact that the comments are specific to the individual student, not generic may encourage student engagement (Shute, 2008). Hepplestone et al. (2011) note that sending feedback comments electronically is an effective means of communicating and can enhance the way in which students engage with it. The fact that the written comments are distributed via a web- based platform ameliorates issues reported with paper based feedback that is never collected (Murdan 2002; Sinclair and Cleland 2007). Accessing the written feedback once it is released is a private matter which students may prefer (Jolly and Boud, 2013). The information can be accessed at a time to suit the learner, many of whom in this context are away from the medical school on elective placements at the time of release. The student can revisit the feedback, perhaps after initial negative emotions have subsided. Although there is no facility for dialogue in this context, the students are embedded in a community of practice (Lave and Wenger,1991) which will facilitate their understanding of the comments in terms of tacit knowledge, jargon or shorthand which may be used.

Watling notes 'feedback is a complex tool for influencing learning' (2014, p692) and '...we cannot approach the use of feedback in any educational setting with the presumption that it will be effective in promoting learning and performance improvement' (2014, p693). The individual written feedback comments provided to the students in the context of this thesis will be of variable quality and utility and may be received differently by individual learners with possibly disparate personal characteristics, motivation and learning goals. At this point it is difficult to predict if

and how the learner and feedback variables may mesh, and if there will be distinct patterns of engagement, depending on the characteristics of both of these.

2.5 Self-assessment of performance

Healthcare professionals are expected to identify their own learning needs through a process of on-going self-assessment. Self-assessment is integral to many formal appraisal processes and is seen as an important aspect of personal and professional behaviour by regulatory bodies. The development of the ability to self-assess is often expressed as learning outcomes for clinical courses of study (Colthart et al. 2008, p124). Sargeant notes that the concept of self-assessment poses 'perplexing challenges' in that it appears self-evident that doctors and other professionals engage in this activity, and yet research which seeks to evaluate the ability to do this accurately questions this assumption (2008, p1).

There are a variety of definitions of self-assessment (Gordon 1991; Boud 1995; Colthart et al. 2008). Epstein et al. define self-assessment as both an externally and internally informed process of interpreting data about one's own performance and comparing it to an explicit or implicit standard (2008, p5). Thus self-assessment is the process by which learners reflect on their abilities and compare where potential gaps lie with regard to their perceived abilities and an external standard. This internally generated feedback can then be compared to externally generated feedback about their abilities. This process of reconciliation of internal and external feedback may also entail an important affective dimension as emotional responses to negative feedback are worked through (Mann et al. 2011). Reconciling the internal and external feedback is integral to sense-making in a social-constructivist learning paradigm and it is only after the feedback is understood and accepted that the learner may be motivated to act on it. The process required for self-assessment is conceptually linked to processes of reflection in and on action (Schön 1991; Epstein 2008; Sargeant et al. 2008; Sargeant et al. 2010; Pelgrim et al. 2013). Self-assessment is also held to be an important element of self-regulated learning activities (Sargeant 2008; Zimmerman 2008; Leggett et al. 2012).

Self-assessment is a complex activity (Epstein et al. 2008; Sergeant et al. 2010). It has been shown that lower performing learners tend to be the least able to accurately self-assess their performance (Gordon 1991; Boud 1995; Davis et al. 2006; Colthart et al. 2008; Boud et al. 2013). The accuracy of self-assessment can be enhanced by providing explicit assessment criteria and benchmarking guidance as well as feedback (Colthart et al. 2008). There is some evidence that it is easier to self-assess practical skills than knowledge-based activities (Colthart et al. 2008). It has been claimed that specific training interventions can improve the accuracy of self-assessments (Srinivasan 2007; Leggett et al. 2012).

Eva et al. state 'while self-assessments may not validly indicate ability, it is still critical to determine how students perceive their ability because their opinions drive their learning goals' (2010, p102). This belief was empirically explored in a study by Bounds et al. (2013). 72 emergency medicine residents (junior doctors in USA) completed a standardised assessment task and were asked to self-assess how they performed via a pro forma. They were then given positive and negative feedback generated from a standardised checklist. The doctors were asked to generate 'SMART' - specific, measurable, attainable, realistic, and time-bound (Doran, 1981) learning goals. The investigators categorised these learning goals as stemming from the doctors' self-assessments, the feedback, or both. It was found that 47% of the total 226 learning goals were generated by self-assessment alone and 27% by the feedback alone. On follow-up, 62 residents recalled 89 learning goals, 52 of which were reported to have been achieved. The goals achieved which were categorised as self-assessment or feedback were equal (25% each) whereas when self-assessment and feedback were in agreement, 40% of the learning goals were achieved. The authors concluded that the alignment of self-assessment with effective feedback gives the greatest effect in motivating actions for improvement.

Self-assessment and this study

In the context of engagement with written feedback comments after an undergraduate clinical examination, I would suggest that the process required for self-assessment is simplified compared to self-assessment in everyday clinical practice, in that the tasks or questions for each station are delineated, so it is clear

to students what they are self-assessing (if they can accurately remember the encounters in the examination). The standard for satisfactory performance for each of the questions is made known to the feedback providers (the examiners) via each question's performance descriptors. Therefore it should be clear what they are judging against when they write the comments under 'what you did well' and 'what you could improve'. However, the student does not see the performance descriptors and their internally generated feedback may not only include the experience of sitting that question; the student may incorporate feedback from other pertinent experiences when they are judging how well they felt they did. The students do not receive any formal training on how to self-assess their performance, so whilst some students may naturally do this, others may not.

2.6 Students' engagement with feedback from clinical examinations

2.6.1. Literature review methods

A structured search of the published literature was conducted in order to provide a descriptive synthesis of what has been published on the topic of student engagement with feedback after clinical examinations. A systematic review of the literature as outlined in Best Evidence Medical Education (Harden et al., 1999) was not conducted for this study as this would constitute a body of research in its own right, which would usually involve a team of contributors. Although systematic reviews are commonly used in health sciences and all fields of education, there is a debate within general education about the appropriateness of these, in part because the related searching and inclusion or exclusion techniques of systematic review are very structured and this restricts the possibility for research questions to evolve (e.g. Hammersley, 2001).

Structured reviews such as this one rely on the author's knowledge and experience to identify similarities in the aims, methodologies or findings in relevant studies, compared to a systematic review or meta-analysis, which uses strict criteria and statistical techniques to examine and compare studies. Although a systematic review was not conducted, sufficient relevant material was examined, defined by Robson as 'relevant works are those that have important implications

for the design, conduct or interpretation of the study, not those that simply deal with the topic ...of the research' (2011 p51).

Four online databases; PubMed, Web of Science, ProQuest and Ovid were searched using the combined search terms ('feedback' OR 'feedforward') AND ('student' OR 'undergraduate') AND 'clinical' AND ('assessment*' OR 'examination*' OR 'OSCE*') on the fields resource titles/abstract, with no date restriction and filtered for English language. The term 'engagement' or possible alternatives were not included because there are so many potential terms that could be used to describe this (e.g. use of, utilisation, up take, effects of, acceptance, act on, read, look at, reflect on, think about) that relevant work might have been overlooked. The terms formative or summative were not included as I felt that this differentiation in the purpose of a clinical examination was not important for this study. The term 'written' was not included as studies describing the use of other 'permanent' feedback formats (e.g. audio or video) would be relevant to include in this study. The terms 'student' and 'undergraduate' were used to limit the search to this stage of training which is the most relevant to this thesis , but not 'medic*' as results from other healthcare professionals were felt to be relevant.

The exclusion criteria were that the papers had to be in a published journal, but not necessarily be peer reviewed. The article had to report on an empirical or review study which included details about the provision of feedback to candidates on their performance in a clinical examination via permanent media and the study participants had to be undergraduate healthcare professionals. There were no methodological exclusion criteria.

After removing duplicate copies of material returned from different databases, this search produced about 90 articles. The titles and abstracts of these were read for relevance to the research question, resulting in a short-list of 30 articles. The reference lists in these articles were reviewed to see if they led to other articles of interest. I later decided to exclude studies which reported on feedback provided by trained 'standardised' patients or by student peer examiners as this made the context of the studies more removed from this study. Feedback from sources

other than a qualified healthcare professional also represented an additional variable that may affect student engagement with feedback, in terms of credibility of the feedback provider. I also decided to exclude studies which were more than 10 years old as they appeared to be too outdated to be relevant (e.g. Black and Harden 1986; Schmidts 2000).

2.6.2 Literature review findings

The final criteria returned seven studies which report on the provision of written, audio-recorded or filmed feedback to undergraduate students during or after a clinical examination. Most of the studies are not overtly framed within any theory of learning, but for some, this can be discerned from the way the study was devised and the results reported. The majority of the studies reported below do not consider whether or how the students engaged with the feedback to improve their knowledge or skills. Instead they evaluate the students' satisfaction with the feedback, or do not consider the students' perspective.

Van Nuland et al. (2012) report on a study in Belgium which focusses on estimating the costs of changing a ten station summative communications skills OSCE to include time for the examiner and the standardised patient to provide written feedback comments, in addition to marks. This study reported that although there was a cost implication of including the two minutes between candidates so the feedback could be written, and having these typed up before being returned to students six weeks later, they felt this was worthwhile. They refer to the 'educational impact' of including the narrative feedback. They administered a survey to students four months after the OSCE which included two questions on the 'educational impact'. The results from 55 responders, suggested that the students appreciated being given written feedback and they felt they learned more from the narrative feedback than the scores. The authors acknowledge that their investigation of the educational impact of the feedback was limited and that they had not explored whether the students really learned from the feedback.

Rush et al. (2014) in the UK report on a novel summative 10 station clinical skills Objective Structured Clinical Assessment (OSCA) for nursing students. Groups of

60-80 students go round seven, five minute long, practical stations in any order they choose until they have undertaken all the procedures within a three hour timeframe. The other three stations are written papers. The students are given immediate feedback from each assessor at the skills stations (mainly via 'tick box' pro formas) to allow them to identify how they could further improve. The student keeps the feedback sheets for their portfolio. They know when they leave the station if they have passed or failed. Students deemed as not competent are offered the opportunity to further practise the skill and then are re-assessed at a later sitting of the OSCA. The students' perspective of this process was evaluated via an on-line 18 question Likert scale survey and seven free text response questions. There were 180 respondents from a total population of 272 students.

Six of the Likert statements related to the perception of the value of the OSCA process, including one item 'Feedback received was valuable' which 98.7% of respondents agreed or somewhat agreed with. A set of four Likert statements explored the positive impact of the OSCE on confidence; clinical skills; knowledge of the theories underpinning the skills; and on motivation to be an excellent nurse. All of these statements had at least 97.1% responses for agree or somewhat agree. One of the themes analysed from the open comments was learning for future practice, relating to comments about how the OSCAs helped students to learn and how they would apply this to future clinical placements. There were a number of positive reflective responses about this aspect. Overall the majority of students felt that the immediate feedback had a positive impact on their reflective skills and understanding of their learning. The authors note that a further study would be beneficial to examine if the results from the OSCA are a predictor of success in the placement-based clinical examinations to provide potential evidence of learning gains. This study considers a holistic approach to evaluating the students' perspectives on pre-assessment learning, the assessment and impact on future learning in clinical practice.

The following set of reports all use computers to mark the OSCE and generate the feedback. Firstly, Ashby et al. (2016) in Australia report on the provision of feedback to occupational therapy and physiotherapy students generated by the use of tablets to electronically mark summative and formative OSCEs (which they

call eOSCEs). The feedback was returned to the student electronically and consisted of a series of performance standards (satisfactory, good etc.) linked to task statements (e.g. good rapport with patient) selected from drop down lists. There was also space to provide free text comments.

Students' perceptions of the utility of the eOSCEs were collected via Likert scale statements on two surveys, (total population N =206); a nine item survey after the formative (respondents N=123); and a 12 item survey after the summative eOSCE, (respondents N=101). The strongest agreement for both surveys was with items on timeliness and ease of access of feedback. There was least agreement with the amount of feedback being sufficient, and for the second survey, if the student considered their performance in the summative eOSCE had improved as a result of the feedback from the formative one. As well as standard items about the timeliness and quality of the feedback comments, there was a question on how the feedback would influence their 'learning for future assessments', rather than phrasing this as for 'learning to improve clinical practice'. The notion that feedback from an assessment is only relevant to that assessment or may only be useful for performance in future assessments suggests to students that healthcare practice and healthcare assessments are two separate, unlinked events. This view resonates with those of the students in the article by Harrison et al. (2014) discussed below.

As part of Ashby et al.'s study, eight focus groups were convened to discuss five open questions about the feedback, including one on what would influence their use of eOSCE feedback in future learning activities. The overall messages from students were that they appreciated the timeliness and accessibility of the feedback which helped them to remember their assessment of how they did at the stations. They made judgements about the examiners' proficiency in observing, marking, and reporting of responses and ability to accurately record the students' responses during the eOSCE. The ability of examiners in handling the technology led to students' confidence in the resulting scores and feedback comments. Thus the students made credibility judgements about the examiners as reported in many of the studies in the section above. The students considered the comments to be most effective when they described how to improve performance and could be

used in future learning, rather than only describing the skills which were incorrectly performed. Interestingly, where free text comments were provided, students reported that they did not always understand the jargon or shorthand used. It is not known how senior the students were, but this finding may have resonance in the context of my research.

Denison et al., (2016) in the UK examined the feedback comments for students from summative and formative OSCEs which were generated by the use of tablet computers compared with those from machine readable mark sheets. In order to do this reliably, they devised a metric based approach, the 'Feedback Quality Rating Scale'. Their findings are that the quality and quantity of comments improved with the use of tablet computers. However, the comments provided are not routinely given directly to all students; instead they are used for feedback purposes during student support interviews. The authors planned to change the relevant box heading from 'comments' to 'candidate feedback' to encourage more comments to be made which could be returned routinely. Students were not asked for their views on the utility of the comments.

Harrison and colleagues provide three reports about the feedback from a UK medical school after summative OSCEs which are marked electronically using tablets. In the first of these listed here, Harrison et al. (2015) describe how students are provided with 'tick-box' feedback for performance strengths and deficiencies in key consultation domains and also with audio feedback which the examiners dictate directly into the tablet. The audio feedback was evaluated via a student survey and focus groups and a staff survey. The student survey had a 65% response rate, with the majority of students accessing the comments and finding them helpful in understanding their strengths and weaknesses; had changed the way they perform a skill; and that they would use the feedback to help them plan for 'future assessments'. The selected comments from focus group members highlighted the contrasting views of students and provided an illustration of the differing expectations of students about the type of feedback they wished to receive. This finding may be very relevant to my study.

The second study listed here, Harrison et al. (2013) is the most pertinent to the context of this study, in that it is a comprehensive empirical study of engagement with feedback. The study describes how feedback after a summative 12 station OSCE is provided via a website designed to display numeric information from the OSCE in a range of ways, and going from global results to detailed comparison pages, by individual station scores, or by skills across stations for individual students, along with graphical whole cohort benchmarking information. There are also a set of 'next steps' pages with guidance on how to interpret, reflect and act on the feedback. Access to the webpages was opened shortly after the results of the OSCE were circulated. The system recorded information about which students looked at specific pages and for how long. This was compared to information about the students' current year OSCE and written paper performance and the previous year's clinical examination. These results were also correlated with data from a 51 item Likert scale survey completed by the students which aimed to measure certain learning related characteristics and attitudes to feedback.

It was found that 96% of 132 students accessed the website, 87% of these on the first day it was available. Analysis of data indicated that comprehensive web page users had a significantly higher OSCE and OSSE score than the minimal users. 'Excellent' students (denoted as those who passed 12 stations) visited the website three or more times, 'just passing' students (those who failed three or four stations) viewed the website the least. 'Excellent' students viewed more global pages than 'failing' students. 'Failing' students used the 'next steps' pages more than 'good' (failed one or two stations) or 'just passing' students. There was no relationship between performance in the written paper and use of the website. With regard to the results of the learner characteristics survey, the comprehensive users scored more highly on the value of feedback scale and minimal users scored more highly on the extrinsic motivation to study scales, but there were no other relationships with learner characteristics.

The authors noted that the students who just passed the OSCE made the least use of the feedback, yet they are at risk of failing future assessments and so potentially would have more to gain from the feedback. They felt further study of

this group was warranted in order to be able to encourage students to aim for optimal capability rather than minimal competence standards. The authors note that for reasons of feasibility they could not include the examiners' written comments in the website and that it is possible that students might have had a different pattern of engagement with narrative comments compared to the numerical feedback. I found Harrison's attention to the differential engagement with feedback by groups of students at different levels of performance attainment very informative and agree with the conclusion drawn that this phenomenon is worthy of further study. Harrison et al. note that a major weakness of their study is that they cannot differentiate between students who simply looked at web pages without attempting to reflect on their feedback, and those who did.

In their third study listed here, Harrison et al. (2014) follow up on some of the themes presented in the previous paper and use a qualitative approach to explore why feedback provided after a summative assessment is not always used effectively by students. Individual interviews were conducted with 17 of the students who had accessed the web-based feedback. Five themes were reported to have emerged from a thematic analysis of the interview transcripts; the power of the summative assessment culture focussed on the need to pass rather than excel, which was a negative influence on the use of feedback; the influence of strong emotions around sitting the examination and receiving the results; the influence of social interactions with others (clinical tutors, peers, parents or partners) reinforcing the notion that passing was sufficient; influence of prior learning experiences on expectations (high attainers in school feeling that they cannot achieve at the same level in medical school and so accept lower goals and stopped trying); disconnect between assessment and future learning so that feedback from the OSCE was not seen as being relevant to clinical practice.

The authors purposely attempted to recruit students from a range of performance levels to ensure that representative views were included in the analysis. However, they did not report if students at different performance levels had similar views on the purpose of feedback or why they chose to the view the web-pages as they did, and what cognitive or affective processes were taking place with regard to the

feedback, although prompt questions on these aspects were included in the interview schedule. I would be interested to find out more about these aspects.

2.7 Conclusion

Much of the literature on feedback on learning focusses on the types of feedback provided, and explores the reasons why students are dissatisfied with it. There is currently a paucity of published reports about the effects on learning associated with feedback from clinical examinations. This chimes with Handley et al. who argue for less research on measuring feedback effectiveness, and more on understanding student *engagement* with feedback (2011, p557). More authors are now suggesting improvements at the meta-level, to ensure that the assessments are planned in order to enhance learning, to provide students with opportunities to develop skills in self-assessment and the feedback literacy skills required to understand why they should, and how they can, engage with the feedback to set actionable learning goals to improve knowledge or skills as required to meet or exceed expected standards.

The brief literature review above provides some insights about common themes which may have an influence in this context. The fact that there is no process of dialogue between student and feedback provider, the students' perception of the 'credibility' of the feedback provider and the influence of negative emotions are all potential influences which may discourage engagement with the feedback in this context. All the influences discussed above should be explored with students to begin to ascertain if and how they use this feedback.

The focussed literature search relating to students' engagement with feedback from clinical examinations suggests that there are a limited number of reports of feedback being provided after clinical examinations via a permanent media, very few of which explore how students engaged with this feedback. There is a gap in our current understanding and therefore it would appear to be a valid endeavour to investigate how the learners in my context engaged with their feedback and if learners at all levels of performance are able to make effective use of their individual feedback comments.

2.8 Summary

The focussed literature review briefly discusses the most commonly referenced theories of learning. Most of these take account of both the learner (their cognitive processes and their personality traits such as motivation and self-efficacy beliefs) and the environment in which they are learning as this influences what is learned and how. If designed effectively, assessment processes can influence how and what is learned and can provide students with effective sources of external feedback. However, the learners must also undertake cognitive and affective processes to self-assess their performance, reconcile their internal and external feedback and be motivated to set learning goals in order to improve. This review briefly explores some of the literature about potential issues with engagement with feedback. There are an increasing number of recent articles which discuss the evaluation of written feedback provided after clinical examinations. This is usually related to the introduction of electronic process of marking which make the collation and dissemination of feedback an easier process to manage. However, in depth investigations into how students engage with this feedback are currently limited and so this would be a worthwhile area to explore.

CHAPTER 3:
UNDERPINNING CONCEPTUAL FRAMEWORK

3.1 Introduction

This chapter will detail the theory of learning which underpins this thesis; self-regulated learning (SRL). I will describe the process by which a conceptual framework was initially devised by combining two existing models that look at the SRL processes from different perspectives and then explain how this outline conceptual framework was further developed by findings from the first empirical study conducted for this thesis. Two surveys were then administered to assess how students engaged with the feedback according to the processes outlined in the conceptual framework.

SRL was chosen as the theoretical approach to conceptualise how students engage with the feedback from their clinical examinations because it is a broad approach to describing cognitive and other academic activities that 'expert learners' undertake (Ertmer and Newby, 1996). It has resonance with several approaches to understanding learning; cognitive, humanist and social learning perspectives, all of which focus on learning as an individual activity. Learners' interactions with their environment and learning with and from others in that environment have important effects. However, ultimately, learning is seen as occurring at the individual level (Mann 2011, p63). Barry Zimmerman (2013), a leading proponent of SRL, describes it as a social-cognitive model, and acknowledges the importance of feedback from others in the learning environment.

Many of the published studies on SRL have school-aged children as their subjects. However, SRL is an appropriate approach for higher education contexts and adult learners (Pintrich, 2004). Cleary et al. (2012) note that there was some debate regarding whether SRL is a stable personality trait or a changeable, teachable skill and provide references to support their contention that SRL skills can be taught and assessed, but caution that this should be done in context rather than

attempting to develop a broad set of skills to be applied to any domain or learning environment.

3.2 Other models of learning

The literature was explored with regard to potential theories of learning that could be adopted for this study. I considered Kolb's (1984) experiential learning theory as much of his work is concerned with the learner's internal cognitive processes, including reflecting on concrete experiences. Kolb stated that effective learning occurs when a student progresses through a cycle of four stages (starting at any point): having a concrete experience followed by the observation of and reflection on that experience which leads to the formation of abstract concepts (analysis) and generalisations (conclusions) which are then used to test hypotheses in future situations, resulting in new experiences. Kolb's theory was linked to his work on 'learning styles' and the potential optimisation of learning environments to improve a learner's experience depending on their learning style preference. I initially considered if students might engage with feedback more or less effectively according to their learning style, but on reflection I thought Kolb's model was too constrained to account for elements of learning such as goal setting, motivation, strategies and choices that are available to, and expected of, students in higher education.

Next I considered Self-determination theory (Ryan and Deci 2000), another cognitive or psychological approach to learning. This theory was clearly elucidated and included some of the same aspects as SRL, such as goal setting. Self-determination theory is focussed on extrinsic and intrinsic motivation in learners, which could potentially be an important influence on the propensity for engagement with feedback. However, SRL appeared to be a more encompassing framework to describe how 'expert learners' (Ertmer and Newby, 1996) use a suite of cognitive and behavioural skills, including the attribute of reflexivity which has been shown to be important in studies about engagement with feedback (Nicol and McFarlane-Dick, 2006; Clark, 2012).

3.3 What is Self-Regulated Learning (SRL)?

Self-regulated learning is a collection of metacognitive skills, habits or processes by which learners are motivated to set and achieve learning goals. Butler and Winne characterised self-regulated learning as:

> '...a style of engaging with tasks in which students exercise a suite of powerful skills: setting goals for upgrading knowledge; deliberating about strategies to select those that balance progress toward goals against unwanted costs; and, as steps are taken and the task evolves, monitoring the accumulating effects of their engagement. ...Self-regulated students are thus aware of qualities of their own knowledge, beliefs, motivation, and cognitive processing'. (1995, p245)

Much has been published about the various facets of SRL behaviour, often with overlapping elements within this set of metacognitive skills being investigated. The work of influential authors on these different elements include; the role of motivation (Zimmerman 2008; Pintrich 2004), differentiating the types of goals that are set by effective self-regulated learners (Dweck and Leggett 1988; Dweck 1999; Grant and Dweck 2003); self-efficacy (Bandura 1977; Schunk 1984; Zimmerman et al. 1992); and self- evaluation (Zimmerman 2008).

Self-regulation has been defined by Zimmerman as 'self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals' (2005, p14). This conceptualisation of learning fits well with the requirement for continuing professional development by practising clinicians, who need to be able to set, reflect on progress and achieve their own learning goals throughout their career. Zimmerman posits that self-regulatory processes and accompanying beliefs fall into three cyclical phases: forethought, performance (or volitional control), and self-reflection processes (2005, p16). These can be summarised as the before, during and after phases of a learning event. Forethought processes by motivated learners include self-efficacy beliefs, goal setting, planning and choice of techniques or strategies. Performance processes occur during the learning activities and include the metacognitive monitoring of learning to ensure that the approaches selected in the forethought stage are the most relevant to ensure that the learning task is successfully completed. The self-

reflective phase includes two facets; 'self-judgement' which includes the processes of self-evaluation (assessing how well the learner did at the task) and causal attribution (the learner's perceptions of the causes of the outcomes they attained in the task which then lead into motivation for the next cycle) and 'self-reaction' which includes self-satisfaction or affect (emotional response) and adaptive/defensive inferences where changes in approach may or may not, be planned for the next learning event (Zimmerman 2008, p178).

However, it has been shown that it is difficult to self-assess ability or performance accurately (Sandars and Cleary 2011, p883; White and Gruppen 2010, p276). Boud et al. (2015, p50) note '…the low ability students significantly overestimated their ability in all their assessment tasks. The high ability group significantly underestimated their grades on all stages of assessments.' However, Eva et al. (2010 p102) state 'while self-assessments may not validly indicate ability, it is still critical to determine how students perceive their ability because their opinions drive their learning goals'.

3.4 The role of reflection in self-regulated learning

> 'As a metacognitive skill which leads to, uses, and subsequently increases metacognitive knowledge, reflection plays a key role in the process of self-regulation. Reflection is critical for transforming the knowledge gained *in* and *on* action into knowledge available *for* action' (Ertmer and Newby 1996, pp18)

The concept and role of reflection in professional learning has been widely discussed in the literature, with a range of definitions provided by different authors.

Nguyen et al. (2014, p1177) stated with regard to medicine, nursing and other health science professionals '...reflection in these contexts is viewed as a crucial component of curriculum and practice, and as a requirement for lifelong personal and professional learning'. Boud et al. (1985) quoted by Sandars (2009, p686) define reflection as '…a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to a new

understanding and appreciation'. Sandars (2009, p685) proposed a wider definition of reflection to include a spectrum of possible uses, approaches and intended outcomes. 'Reflection is a metacognitive process that occurs before, during and after situations with the purpose of developing greater understanding of both the self and the situation so that future encounters with the situation are informed from previous encounters'.

After identifying, exploring and analysing what they felt were the most important conceptualisations of reflection, Nguyen et al. (2014, p1182) defined reflection as

> 'The process of engaging the self in attentive, critical, exploratory and iterative interactions with one's thoughts and actions, and underlying conceptual frame, with a view to changing them and with a view on the change itself'.

Boud and Walker highlight the role of emotions in this process of reflection.

> 'Because emotions and feelings are often downplayed in educational settings, it is common for reflection to be treated as if it were an intellectual exercise – a simple matter of thinking rigorously. However, reflection is not a solely cognitive process: emotions are central to all learning' (1998 p194).

Sargeant et al. conclude that 'Reflection appeared to be the process through which feedback was or was not assimilated and appeared integral to decisions to accept and use the feedback' (2009, p399).

Therefore within the suite of metacognitive skills practised by effective self-regulated learners, the ability to reflect on a learning related event the student has experienced and self-assess how well the initial learning goals were achieved is important. In the context of this study, after the clinical examination, an effective self-regulated learner would reflect on their performance in each of the stations (questions), self-assess and come to a judgement about how well they were able to demonstrate the knowledge or skills expected by the examiner and if they felt they had any deficiencies. They may also act to remediate these deficiencies. Furthermore, based on the findings reported in Chapter 5, if students are internally

or externally motivated to access the written feedback provided after the examination, reflection appears to be one of the key actions required to reconcile the internally generated and externally provided feedback, and decide whether to accept the feedback and subsequently act on it.

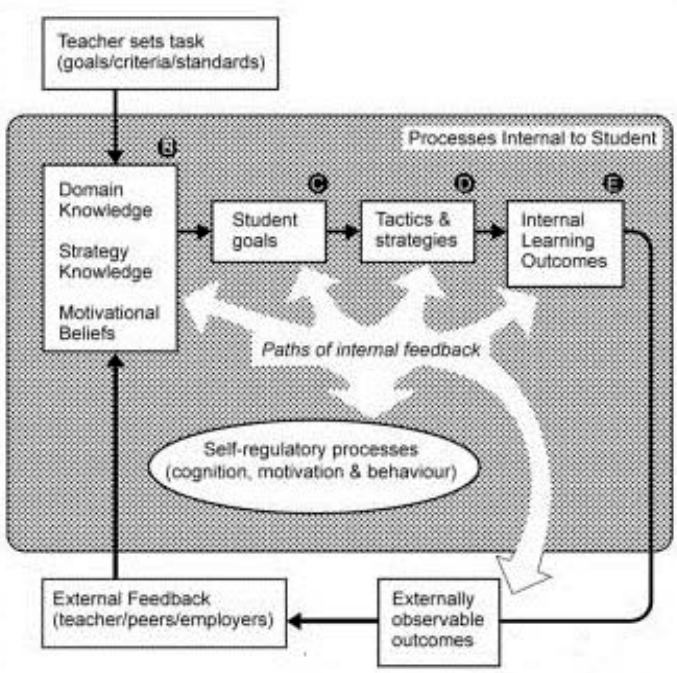
3.5 Adaptation of existing models of SRL to this context

Butler and Winne explain engagement with feedback from a self-regulated learning theory perspective:

> '...learners' knowledge, beliefs, and thinking jointly mediate the effects of externally provided feedback. This mediation is the funnelling through monitoring of information about various topics - task, self, epistemological characteristics of knowledge, goals, and cognitive tactics and strategies - to confirm, overwrite, add to, tune, or restructure extant knowledge and beliefs'. (1995, p275)

Butler and Winne provided a model of self-regulated learning (1995, p248) which Nicol and Macfarlane-Dick (2006, p202) updated to emphasise the importance of providing well-constructed feedback to support and develop learner self-regulation (Figure 3.1). Nicol and Macfarlane-Dick's model specifically related to feedback after formative assessments. They quote Sadler's definition 'Formative assessment refers to assessment that is specifically intended to generate feedback on performance to improve and accelerate learning' (Sadler 1998, p199). Taras argues that 'all assessment begins with summative assessment (which is a judgement) and that formative assessment is in fact summative assessment plus feedback which is used by the learner' (2005, p466). With regard to the MB ChB programme at Birmingham, in addition to marks, various types of feedback are provided after all summative assessments. Therefore, the distinction between feedback provided after summative and formative assessments is not relevant in this context. However, as discussed in Chapter 5, learners at lower levels of performance do not necessarily appreciate that feedback from a summative assessment is intended as feed forward prompts to improve future clinical practice.

Figure 3.1: Nicol and Macfarlane-Dick's (2006) Conceptual model of processes of self-regulated learning

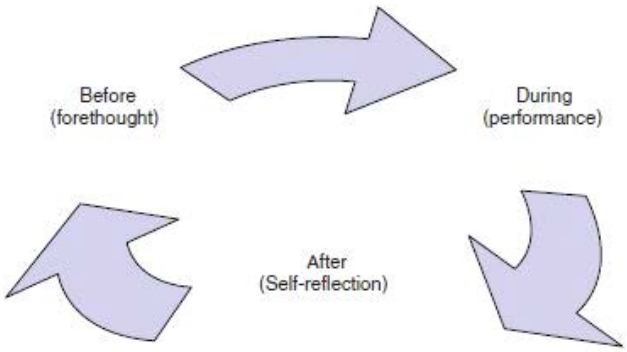


Previous studies have highlighted the crucial role of reflection in effective self-assessment of progress and subsequent engagement with external feedback (e.g. Sargeant et al. 2009, White et al. 2009, Pelgrim et al. 2013). Nicol and Macfarlane-Dick did include reference to reflection, but called it 'self-assessment' (2006, p.205). Facilitation of student self-assessment was included as the second of their 'seven principles of good feedback practice', so in their model reflection was shown as a process for potential external development, rather than a current activity, internal to the student.

Sandars and Cleary (2011) proposed a model to illustrate their three phases of self-regulation. This did not explicitly relate to an assessed activity or the incorporation of external feedback after the activity, but they note the importance of the 'after' or 'self-reflection' phase which was also emphasised by Zimmerman (2005). This is an important step which effective self-regulated learners could equally be expected to undertake after completing an assessment activity. If effective self-regulated learners are believed to incorporate reflective internal monitoring or self-feedback on their progress towards their learning goals, it could be assumed that they would also have the cognitive routines to self-assess their

performance after (and possibly even during) an assessment of that learning and so this was added to the outline conceptual framework shown in Figure 3.2 below.

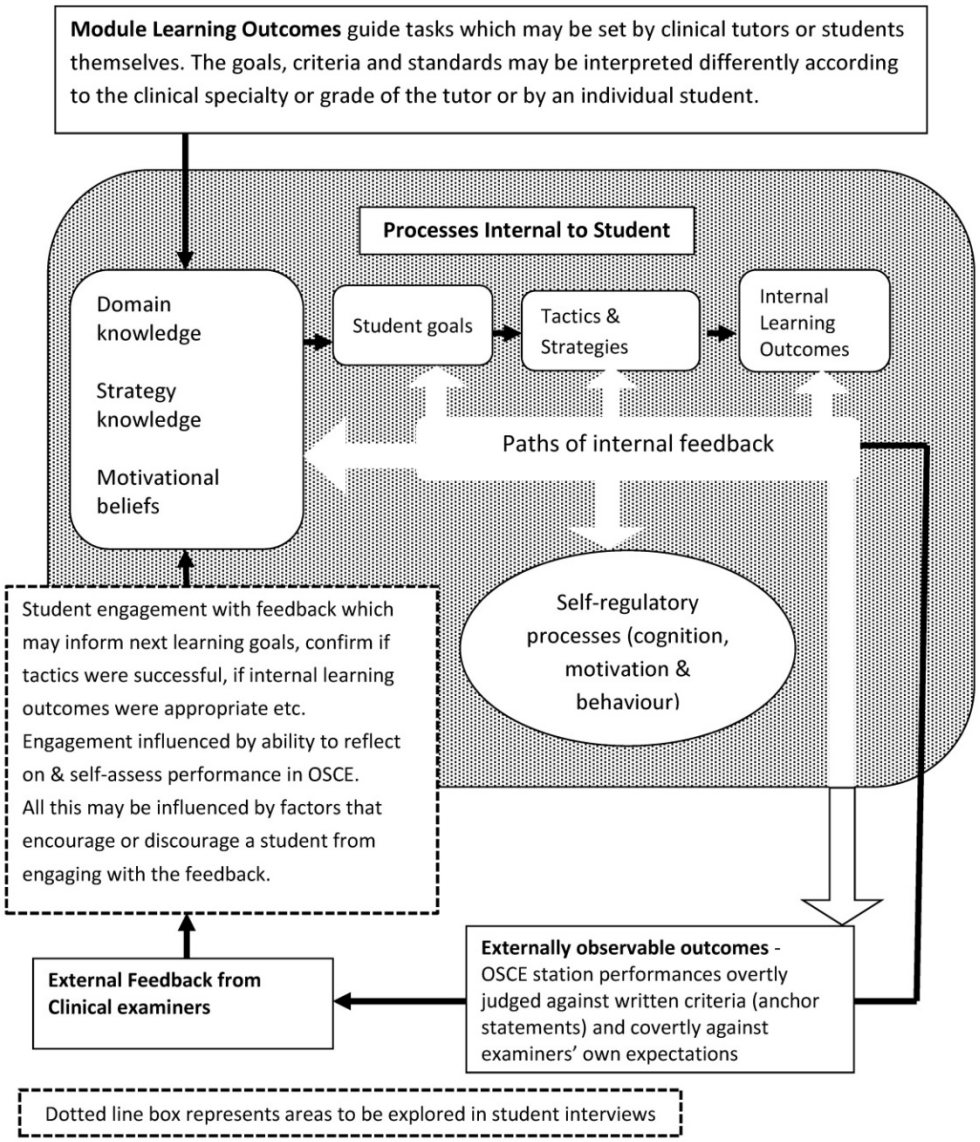
Figure 3.2: Sandars and Cleary's (2011) Phases of self-regulatory processes



As discussed in Chapter 2 and Chapter 5, various studies have highlighted different barriers to engagement with feedback (Bing-You et al. 1997; White 2007; Sargeant et al. 2008; Sandars and Cleary 2011; Watling and Lingard 2012; Agiusa and Wilkinson 2014; Patel et al. 2015). Nicol and Macfarlane-Dick's conceptual model did not include reference to the fact that, for a range of reasons, feedback might not be accessed, accepted or acted on, so this was also added to the initial outline conceptual framework (Figure 3.3).

I thought that performance level in the clinical examinations might be a potential proxy measure of students' ability in effective self-regulation strategies, in that effective self-regulated learners would achieve better scores in individual questions and the examination overall (Sundre and Kitsantas 2004). Therefore for the first empirical study described in Chapter 5, it was planned to recruit a purposeful sample of respondents, stratified by performance level in order to explore if students at extremes of performance level reported different views or actions about their engagement with the feedback, which would then influence the further development of the outline conceptual framework (Figure 3.3).

Figure 3.3: Initial outline conceptual framework of processes in self-regulated learning incorporating engagement with feedback contextualised to Y4 MB ChB at Birmingham Based on Nicol and Macfarlane-Dick (2006) & Sandars and Cleary (2011)



3.6 Further development of the SRL conceptual framework

In the light of responses to the interview questions which are discussed in detail in Chapter 5, the initial conceptual framework above was further developed by adding details about the different thoughts and actions that the students reported they undertook. For example what they thought and did *after* the examination and *before* the feedback was released. This step in the development of the conceptual framework is shown in Chapter 5, Figure 5.2.

The completed conceptual framework is shown in Chapter 5 as Figure 5.3. This is contextualised to the clinical examination held at the end of the fourth year of study as an example, but the processes are exactly the same for the third and fifth years which have eight and 18 stations or questions respectively in their clinical examinations. Two survey studies were then devised in order to measure students' engagement with the individual feedback comments via the steps presented in the conceptual framework and to explore if there was an association between different components of engagement conceptualised as a hierarchy of potential thoughts and actions and the respondents' scores from the clinical examination. The conceptual framework also highlights the importance of the ability of the students to reflect on their performance in the examination and if there is congruence between this and their feedback comments. Sub-scales from a previously validated survey devised to assess two levels of reflection on learning on a course of study were also included in the two survey instruments. The process for selecting the instrument to assess reflection is discussed in Chapter 6.

The first survey study (Phase 1) described in Chapter 7 attempted to assess students' hierarchical levels of engagement with the individual feedback comments via items allocated to sub-scales described as 'Reads', 'Thinks' and 'Acts' and to explore if there was an association between these survey sub-scales and the respondents' scores from the clinical examination. In the second survey study (Phase 2), which is described in Chapter 8, these initial sub-scales were updated in the light of the results of the first survey study. The conceptual framework was not changed after the Phase 1 or Phase 2 survey studies as it was felt that it did encompass all the expected elements of effective engagement, it was the survey instruments that attempted to assess these that required change.

3.7 Discussion

I postulated that the framing of Zimmerman's (2005) cyclical process of SRL could be extended to clinical examinations with the 'before' phase being the self-regulated activities pertaining to learning or revising specific topics, the 'during' phase being the active demonstration of knowledge and skills in the examination and the 'after' phase being when the student reflects on what and how they did or said during the examination. This after phase involves self-assessment or

generation of internal feedback of what the student was expected to demonstrate to the examiner against how well the student thought they did and may also be benchmarked against the student's prior learning goals or performance expectations.

As shown in the finalised conceptual framework in Chapter 5, Figure 5.3, effective self-regulated learners also have the motivation to subsequently access the externally generated feedback provided by the examiners and the metacognitive skills to critically reflect on the internally generated and externally provided feedback in order to reconcile any differences between these and, as felt appropriate, act to improve deficits in learning or knowledge after this process.

Previous research has shown that effective self-regulation is a distinguishing characteristic between high and low performing students in HE settings (Sundre and Kitsantas 2004). However, Sandars cautions '…learners may be highly self-regulated in one context, but not in another. Like most academic skills, self-regulated learning is context-specific' (2013, p1162). Brydges and Butler agree 'However, an individual's approach to self-regulation is not a direct reflection of context. Individuals bring to contexts a variety of knowledge, beliefs and emotions that have developed over time through their history and experiences that emerge in particular settings' (2012, p 74).

Students may not only vary in their overall approach to self-regulated learning depending on the context of the learning or assessment, they may have more or less effective self-regulation skills with regard to each of Zimmerman's three phases. Students may be less effective at the self-reflection phase after an examination in terms of how they self-assess; the accuracy of their self-assessment compared with the examiner's expert judgement; and the conclusions students draw about their performance in the exam. All students will access the summative feedback from the exams i.e. the marks, because of the consequences which follow passing or failing the assessment. Passing students may not be sufficiently motivated to access the formative feedback on their performance, or if they do read it, to reflect critically on it and take internally motivated action to improve. Failing students have external motivation to access the feedback and

actively attempt to improve their knowledge or skills before the next diet of the examination.

3.8 Conclusion

The initial outline conceptual framework encapsulated the expected processes that an effective self–regulated learner would undertake with regards to the initial learning or revision during a course of study and the engagement with feedback generated after the summative assessments at the end of the course, which would then inform and motivate the learner's decisions about what and how to learn in the next cycle of study. These processes were explored in an interview study with participants who were stratified by performance in the end of Y4 clinical examination. It was found that the outline conceptual framework included the processes, but not all of the steps in these processes. Therefore it was revised to include the step of reflecting on performance before feedback was provided as well as after accessing the feedback. The finalised SRL conceptual framework was then used to inform the development of a survey instrument to assess the respondents' hierarchical levels of engagement with the individual feedback comments. The conceptual framework appeared to be fit for purpose and no further changes were made after the Engagement with Feedback survey was administered over two phases.

3.9 Summary

This chapter briefly explains why SRL was selected as the theoretical underpinning for this thesis study and details Zimmerman's (2005) explanation of SRL as a cyclical process. I explain why this is an appropriate representation of actions relating to engagement with feedback. The importance to this study of reflection as a facet of SRL is examined. I introduce an initial outline conceptual model of SRL contextualised to this study and document its development by incorporating findings from the first empirical study I undertook. I explain how the finalised conceptual framework was then used to devise items for survey sub-scales to assess engagement with feedback which were administered along with sub-scales from an established survey to assess reflection on learning during a course of study.

CHAPTER 4: METHODOLOGY

4.1 Introduction

The theoretical framework underpinning any research informs the basic assumptions that guide the research and encompasses the elements of ontology, epistemology and methodology. There are a number of research paradigms or what Creswell (2014, p5) calls 'worldviews' which is the researcher's set of beliefs or assumptions about what is permissible or valid to be studied, how it should be studied and how the results should be interpreted. This chapter will present my position on these issues. It will explain the overarching research design and discuss in detail the methods used to generate and analyse the data presented in this thesis in order to provide evidence to support the credibility of the findings presented in the following chapters.

4.2 Ontology and Epistemology

Ontology is the study of being or thinking about the nature of existence and the nature of social reality. Epistemology is the theory of knowledge, its evolving nature and the limits to existing knowledge. There has been much discussion about the ontology and epistemology of research via qualitative methods conducted in the social sciences and if and how this should be compared to the ontology and epistemology of research in the natural sciences. However, it is now generally agreed that promulgating this dichotomy of approaches to thinking about what is knowledge and if certain research methods should have primacy, is an unproductive discourse (Bryman 2006, p111).

The debate between exponents of the two opposite paradigms; positivism and interpretivism has enabled some convergence of these two positions. The paradigm shift from positivism to post-positivism acknowledges that researchers in the physical sciences are not disembodied automatons, but are humans in the real world and their life experiences and positionality influence their decisions and actions (Robson, 2011, p22). There has also been growing acknowledgement from researchers within the post-positivist paradigm that the controlled experiment method of 'discovering' knowledge is fallible: 'More than 70% of researchers have

tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments' (Baker, 2016).

Similarly, it has been highlighted that researchers following an interpretivist paradigm in the social sciences may not necessarily seek to generate knowledge which is generalisable but they should follow a systematic approach, and be explicit about how they generate and analyse data and their role as a person making interpretations about the thoughts or actions of other people (Robson, 2011, p15).

4.2.1 Philosophical approach to this research

Research in medical sciences is generally practised in a positivist/post positivist framework, with the randomised controlled experimental approach being seen as the 'gold standard' and results usually exclusively reported via quantitative methods. Research in medical education generally followed this paradigm because this is what doctors have experience of and understand as being the way to 'determine the truth,' and so applied these methods to educational research practices. However, there is increasing agreement in the medical education literature that a paradigm shift in the research methodologies used would be beneficial, for example Regehr (2010). There are also articles which explain and compare different potential approaches to health education research, for example Bunniss and Kelly (2010), and Ringsted, Hodges and Scherpbier (2011).

I was initially drawn towards post-positivist, experimental methods because this is the approach commonly understood by my colleagues in the medical and biosciences disciplines. However, since reading the wider educational literature, I also understand the potential veracity of interpretivist (or constructivist), qualitative methodological approaches in this context. I would characterise myself as having a pragmatic approach to education research. Robson (2011, p28) provides a list of some features of the pragmatic approach, a sub set of which I have replicated below to illustrate this approach and to represent my position.

The pragmatic approach:

- Seeks a middle ground between philosophical dogmatisms and scepticism.
- Rejects traditional dualisms (e.g. rationalism vs. empiricism, facts vs. values).
- Recognises the existence and importance of the natural or physical world as well as the emergent social and psychological world.
- Places high regard for the reality of, and influence of, the inner world of human experience in action.
- Views knowledge as both being constructed and based on the reality of the world we experience and live in.
- Endorses fallibilism (current beliefs and research conclusions are rarely, if ever viewed as perfect, certain or absolute).
- Endorses a strong and practical empiricism as the path to determine what works.
- Endorses theory that informs effective practice.

4.3 Research Design

The research data presented in this thesis were generated from both quantitative and qualitative methods as this mixed methods approach was felt to be the most appropriate to inform the evolving research questions. The research design was an exploratory sequential mixed methods design whereby 'the researcher first begins by exploring with qualitative data and analysis and then uses the findings in a second quantitative phase' (Creswell 2014, p225). The initial qualitative interview study sought to assess, and inform the development of, an outline conceptual framework for self-regulated learning. The interview study enabled the identification of key ideas that could be tested or measured using subsequent quantitative approaches to analysing data generated via survey instruments.

4.4 Research Methods - Semi-structured interviews

Ringsted et al. note (2011, p699) that

> 'Theoretically, qualitative research arises from a twentieth century development in philosophy that recognised the importance of language in constructing what we perceive as 'reality'. Qualititative approaches are best

used for discovering the answers to 'why', 'how' or 'what is the nature of…' type questions.'

The semi-structured interview is an established method to garner data about a subject of interest from people who are directly involved in that subject. Kvale states that 'The qualitative research interview attempts to understand the world from the subjects' points of view, to unfold the meaning of peoples' experiences' (1996, p1).

4.4.1 Development of semi-structured interview schedule

The review of the literature on feedback in educational settings and self-regulated learning (SRL) which is reported in Chapters 2 and 3 informed the development of an outline conceptual framework of student potential engagement with individual feedback comments, including reasons why students might not engage. The questions asked in the semi-structured interviews were devised to firstly assess if the initial conceptual framework effectively outlined students' potential engagement with the feedback comments in terms of accessing the comments, cognitively engaging with them and acting on them to improve knowledge or skills. Secondly, if so, to inform the development of this framework by providing details about exactly what students at different levels of performance in the clinical examination thought about and did with the feedback.

The initial questions on the interview schedule were included to open the interview and build rapport with the students. These questions were more general and neutral (from my perspective), to get students thinking about how they approached their learning activities, in order that I could understand more about them as students and gain a wider perspective on their learning proclivities. The questions then focussed on students' self-assessment of performance in the OSCE and their actions before the feedback was released and feelings, thoughts and actions with regard to the feedback comments once these were available. I did not have the time to be able to pilot the questions as suggested in Braun and Clarke (2013, p81). However, I did ensure that the questions were 'organised so that they flowed logically' (p84) and would enable me to answer the research questions (p85). The interview questions are included as Appendix 2.

4.4.2 The interviews

I decided not to follow an 'iterative process' and use concurrent data analysis to inform data collection (Kuper et al. 2008a) by overtly changing the interview questions in the light of responses from previous respondents because I was assessing an existing theoretical framework, not attempting to generate one. Pragmatically, it would have added a layer of scheduling complexity that might have deterred some potential respondents. The interviews were scheduled for mutual convenience, usually in the evening and I did not know when I agreed the first interview date in what order or how many students from each of the performance groups would volunteer.

4.4.3 Data Analysis

The participant information circulated with the email inviting students to attend for an interview included the fact that the interviews would be audio taped. Braun and Clarke (2013, p92) note this is beneficial because the 'richness and detail' of the interview will not be lost. In order to maintain confidentiality, directly after each interview the audio file was transferred to my personal, password protected area on the secure University server and deleted the file from the recording machine, which was on loan. In order to maintain anonymity, student names were not included in any transcriptions, instead their responses were anonymised by being given a code comprising of the order they were interviewed in, followed by their performance level (T= top; B= borderline; R= resit) and their gender. I transcribed the interview discussions as soon as possible after each interview. Although I am not an efficient typist, this approach was taken so that I could begin to 'immerse' myself in the data (Braun and Clarke, 2013, p204). The transcription of each discussion was almost verbatim so that there was a complete record to subsequently analyse. Non-language communication (e.g. laughter, pauses or 'erms') was not included because it was not relevant for this particular type of analysis. The full details of digressions into technical aspects of the marks and standard setting processes were not transcribed as these were not directly relevant to this study. Instead, a note was included about the general discussion. As recommended in Braun and Clarke (2013, p168), a few days after each transcript was finished, I re-played the audio recording and compared this to the transcript to check the accuracy of the written data. Where necessary, changes

were made to the transcript to align them with the audio. I could not interpret a few of the words in each interview, even after repeated listening.

The typed transcripts were then printed out and read and sections of particular interest were highlighted, to facilitate the process of beginning to consider common responses and emerging themes. To facilitate further analysis of the data, each transcript was then imported into QSR NVivo 10 software. Saldana (2013, p3) states 'A code in qualitative enquiry is most often a word or short phrase that symbolically assigns a summative, salient, essence capturing and/ or evocative for a portion of language based or visual data' and 'a theme is an outcome of coding, categorisation or analytic reflection, not something that is in itself, coded' (p14). A thematic coding analysis as described by Robson (2011, p476) was conducted. This entailed familiarisation with the data (by listening to the discussion whilst transcribing, reading the transcripts and noting down initial ideas), generating initial codes, identifying themes, constructing thematic networks and then integrating and interpreting the resulting patterns.

The codes I devised ranged from the affective (emotions and attitudes) to the descriptive or contextual (e.g. mentions of effects of exam content or format) and as is recommended, much of the data was coded to two or more categories (Braun and Clarke, 2013, p207). In keeping with a constructivist methodological approach appropriate to this type of study, I wanted to take cognisance of any responses that might not appear relevant to the initial research questions, but might hint at other issues that were of concern to the respondents and may have impacted on the efficacy of the proposed self-regulated learning conceptual framework. Therefore, almost all of the dialogue was coded in what Braun and Clarke (2013, p206) refer to as a 'complete coding' approach. A second coding cycle was then undertaken to conflate the initial more specific codes into themes and to synthesise the data in order to answer the research questions.

4.5 Research methods - Survey instruments

Creswell (2014, p13) describes survey research as providing 'a quantitative or numeric description of trends, attitudes or opinions of a population by studying a sample of that population'. He notes that the sample may be cross sectional or

longitudinal and that surveys may be used to collect data with the intent of generalising the findings from the sample to a wider population. In order to assess and further develop the conceptual framework described in Chapter 3, I decided to administer a questionnaire to volunteer student respondents that would aim to measure both their engagement with the feedback and their skills at reflection. The rationale for selecting the Reflection Questionnaire (Kember et al., 2000) and using two of the sub-scales from this is described in Chapter 6.

The questionnaire included basic demographic questions in order to be able to assess how well the sample of responders represented the whole population and a question asking for the score achieved in the end of year clinical examination in the previous year in order to be able to explore if the scores in the OSCE was associated with the scores on the scales to assess engagement with feedback and/or reflection on learning.

The Phase 1 questionnaire was paper based and designed in Microsoft Word to be scanned by an optical mark reader (OMR) machine and so respondents were asked to fill in small circles to indicate their answers. Students complete similar pro formas for their answers in multiple choice question examinations and so are used to this format. Two free text comment boxes were also included; one for feedback from respondents on difficulties they encountered in understanding or completing any of the demographic and OSCE results questions; and one box after the 20 Likert items to report any issues encountered with understanding or answering the questions in that section. It was expected that the questionnaire would take approximately ten minutes to complete. The Phase 1 questionnaire is included as Appendix 3

The Phase 2 questionnaire was very similar in layout and content to the Phase 1 version, but was also made available to certain groups of students in an electronic format via the Bristol Online Survey tool in order to be able to gather data from a wider sample of students. There were minor changes to the wording of the demographic questions (fully explained in Chapters 7 and 8) following feedback comments from Phase 1. No space was included for the provision of further feedback comments. The Phase 2 questionnaire is included as Appendix 4.

4.5.1 Development of Phase 1 'Engagement with Feedback' items

A number of scales that purport to measure feedback seeking behaviour by employees have been published (e.g. Ashford, 1986; Fedor et al., 1992; Linderbaum and Levy, 2011). I reviewed these scales and decided that the most appropriate one for the aims of this thesis was the Feedback Orientation scale (Linderbaum and Levy, 2011). I considered if, after adapting the wording of most of the items (as they related feedback in work-places), some or all of the sub-scales in the Feedback Orientation scale could be used. However, after further consideration, I felt that although the authors of this study presented a strong case for the potential validity and reliability of this scale, I did not want to measure students' orientation to feedback at the abstract level as I was planning to do for reflection. Instead, in order to assess the conceptual framework described in Chapters 3 and 5, I wanted to ask contextualised questions in order to find out more about if and how students engaged with their personalised feedback comments. Therefore I devised items to produce a scale to assess 'Engagement with Feedback' (EWF). Following the conceptual framework, the scale was comprised of three sub-scales, 'Reads', 'Thinks' and 'Acts'. 'Reads' relates to the students' propensity to read the feedback comments; 'Thinks' relates to the students' propensity to interact cognitively with the feedback, judge the comments against what they previously thought of their performance and agree or disagree with the comments. The students may also have an emotional response to the feedback. The students' initial thoughts and feelings about the feedback may change with further reflection; 'Acts' relates to the students' propensity to do something to improve their knowledge or skills.

I purposely constructed a small number of items, unlike the approach usually described in the literature (e.g. Pallant 2013, p188) where one is advised to start with larger number of questions to administer as a pilot and perform factor analysis techniques to ascertain which statements were the most appropriate and delete the more poorly performing items. This was because I had a 'detailed conception of the target construct and its theoretical context' (Clark and Watson 1995, p310), drawn from the previously constructed SRL conceptual framework and the earlier interview study. I showed the items to a small group of junior doctors and asked

them if they had any suggested additions or amendments to the items and they did not.

I included some positively worded and some negatively worded statements as recommended by Robson (2011, p304) 'Items should reflect both a positive and a negative stance to the issue'. In order to reduce potential 'social desirability' response bias, or responding in a way that shows the respondent in a good light (Robson 2011, p240), I included a plausible reason for not thinking or acting with the reverse scored items 9, 17 and 20 (the survey items are shown in Appendix 3).

I decided to use the same number of scale descriptors in the administration of the 'Engagement with feedback' items as the original Reflection questionnaire. These were 'definitely agree', 'agree with reservation', 'only to be used if a definite answer is not possible', 'disagree with reservation' and 'definitely agree' for fidelity in using their questionnaire and for consistency across the questionnaire, although the nomenclature for the three middle categories was not particularly intuitive. Both these scales were scored from 'definitely agree' scored as five to 'definitely disagree' scored as one. As recommended by Hartley and Betts (2010, p25) the scales have the 'strongly agree' column on the left.

4.5.2 Development of Phase 2 'Engagement with Feedback' items

In the light of the findings from the factor analysis of the Phase 1 data, the three negatively scored items were replaced with positively worded statements, designed so that the three EWF sub-scales would each have at least three items (Table 4.1).

Table 4.1: Items changed between questionnaire administrations

	Phase 1	Phase 2	Sub-scale
Q9	I did not think about how I had done on each station because it was too late to change anything.	I thought about the knowledge and skills I was asked to demonstrate at each station.	TPF
Q17	As I wanted to start this year fresh I am not planning to act on the suggested areas to improve in my individual feedback comments.	I read my individual feedback comments to find out about my knowledge and skills.	LLE
Q20	Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them.	It is important to act on my individual feedback comments to improve my knowledge and skills in this academic year.	HLE

NOTE: Scales TPF = Thinking for pass/fail; LLE = Lower level engagement; HLE = Higher level engagement are described in Chapters 7 and 8

4.5.3 Data analysis

Analysis of Likert scale data

Jamieson (2004, p1217) questioned the practice of treating Likert scale generated data as interval type and using parametric statistical methods to analyse this when in reality it is ordinal data for which non parametric methods should be used. However, Carifio and Perla, 2008 responded that data from collections of items from Likert scales are interval, compared to individual Likert items which are ordinal and it is common practice to treat the composite data as interval type.

Pallant (2013, p129) notes 'Spearman's rho correlation co-efficient used for ordinal or ranked data is increasingly being used in psychology research as researchers become more aware of the potential problems of assuming that ordinal level ratings (e.g. Likert scales) approximate interval level scaling'. However, for the analyses of the data generated via the two questionnaires, I decided to follow the methods used by authors of studies related to the area of interest in this thesis and use parametric statistical methods.

Exploratory Factor Analysis (EFA)

In order to assess if the initial, hypothetical sub-scales or factors; Reads, Thinks and Acts, described above did represent the intended hierarchical approach to characterising engagement with feedback, the data from the answers to the survey

were subject to exploratory factor analysis. This is a method of exploring the data to show if there are valid relationships between the items (within and between sub-scales) and so to assess the conceptual framework. A software package, IBM SPSS Statistics version 22 was used to conduct all the statistical analyses.

There are various methods available for identifying (extracting) the number of underlying factors or dimensions in a data set from a survey and Pallant states that the most commonly used approach is principal components analysis or PCA (2013, p190). Pett et al. (2003, p102) note that;

> 'PCA is a straightforward, easily understood and commonly used extraction technique in factor analysis. Its goal is to arrive at a succinct set of uncorrelated components that extract variance in descending order and that can empirically, parsimoniously, and effectively summarize the data set'.

There are different statistical tests to assist in the decision about the number of factors to further analyse. Costello and Osborne caution that 'EFA is a complex procedure with few absolute guidelines and many options' (2005, p1). Therefore, I planned to use the three approaches described below as appropriate to investigate which might best inform decisions about the potential number of factors to retain. Kaiser's criterion rule states that only factors with an eigenvalue of 1.0 or more should be retained for further analysis. The eigenvalue of a factor represents the amount of the total variance explained by that factor (Pallant 2013, p191). Catell's scree test can also be used to inspect the eigenvalues (Catell, 1966). The eigenvalues of all the factors are plotted and the point at which the shape of the curve changes to become more horizontal represents the number of factors to retain (above this break point). Horn's parallel analysis is another method which can be used to decide the most appropriate number of factors to retain (Horn, 1965). Parallel analysis involves comparing the eigenvalue scores with those obtained from a randomly generated data set of the same size. The factors with eigenvalues that are higher than the corresponding random data set are included for further analysis. There are a number of free random data generating software programme available and Pallant recommends the Monte Carlo programme (2013, p.200). However, after conducting these statistical tests,

it should be noted 'whatever solution we arrive at should not be solely based on statistical criteria; it also needs to make theoretical sense'. (Pett et al.2003, p116).

Choice of rotation method

Once the appropriate number of factors has been decided, the factors can be 'rotated' by the software package so that the questions included in each factor have the maximum amount of shared variance. The software programme does not label the scales or interpret their meaning; it presents the patterns of variance loading in descending order per factor so it is easier to elucidate the results with reference to an underpinning theory. There are two main methods of rotation which produce uncorrelated (orthogonal e.g. Varimax method) or correlated (oblique e.g. Direct Oblimin method) factor solutions. I assumed that the underpinning concepts which the factors from the data would represent would be correlated as the three sub-scales represent a continuum of engagement so planned to perform the Direct Oblimin method. Costello and Osborne note 'In the social sciences we generally expect some correlation among factors, since behaviour is rarely partitioned into neatly packaged units that function independently of one another' (2005, p3).

Pallant advises 'I always recommend starting with an oblique rotation to check the degree of correlation between your factors' (2013, p192), although de Vaus (2014, p.195) suggests that Varimax is the normal method to use. In order to check some of the results of the factor analyses, I did conduct some orthogonal (Varimax) analyses and compare the results which were similar.

4.6 Ethical Issues

Ethical approval for each of the studies was granted by the University of Birmingham's Humanities and Social Sciences Ethical Review Committee following the usual processes (Reference numbers ERN_14-0855 and ERN_16-0010). Permission to contact the students was granted by the Programme Director. There was no external sponsorship or any other funding to conduct the studies.

The British Educational Research Association (BERA) provides ethical guidelines for educational research (2011). These highlight a number of ethical considerations to take into account in all research designs with regards to a researcher's responsibilities to participants. Those that are most pertinent to these three studies are listed below.

Firstly, it is important that participants voluntarily give their consent before the research commences. To be able to give informed consent, participants must be given a full explanation regarding the study, what its purpose and aims are, what the methods of data collection are, what the results will be used for and how they will be reported or disseminated. Secondly, participants need to know what their time commitment will be (including the time to fill in any questionnaires or to provide personal information or views). Thirdly, participants need to be given information about their right to withdraw their consent (and their data), the process by which they can do that (with any deadlines by which it will be feasible to do this clearly specified) and the fact that there will be no consequences to them if they do withdraw. Fourthly, it is important that participants are made aware of the steps that will be taken to protect the confidentiality of their data, for example how it will be stored, who will have access to it and for what purposes and when the primary data will be destroyed and how. Fifthly, any predictable detriment arising from the process of findings of the research should be made known to the participants. All the information relating to the above points was provided for each recipient and is shown in Appendix 5.

With regard the interview study, confidentiality was maintained by storing the interview information and the data from the surveys in a password-protected file on a secure internal University server. All paper printouts of interview transcripts were mechanically shredded immediately after the data had been analysed. With regard to protecting the anonymity of interview participants, they were referred to via a unique identifier in data analysis and reporting. The key to their names and contact details was kept in a separate folder on the secure server. For the survey studies, the first phase was completely anonymous. For the second phase, students were given the option of providing their University registration number so that their demographic and OSCE score data could be linked afterwards. In order

to maintain confidentiality, this was not done by the researcher but by the College of Medical and Dental Sciences Data Manager, who routinely has access to the data, and the resulting merged data was returned with a unique identity number. All data for this thesis will be retained for ten years following completion of this study and will then be electronically erased or physically destroyed, as per University of Birmingham guidelines.

Robson (2011, p218) raises the issue of knowing what to do if you discover something you know or suspect to be illegal, or other situations that may be a cause for concern. If an interview participant had revealed something about their professional behaviour and attitudes that brought into question their potential fitness to practice, then it might have been appropriate to break confidentiality. There is an established MB ChB programme organisational structure and process to report concerns, which is widely published to all students.

It is not obvious how there could have been any detriment to the participants taking part in this study, apart from the opportunity cost of the use of their time. There was no coercion and no incentives were offered to participants apart from light refreshments during the interviews. Participants were Higher Education students, so were intellectually equipped to understand what the research is about. The participants from the later years of the course may have conducted their own bio-medical or educational research projects, so would be familiar with the processes. The interviews and requests for the survey instruments to be completed were conducted in meeting or lecture rooms in the Medical School, so there were no personal safety issues for the researcher or responders to consider. There may have been an ethical issue relating to the potential wasting of respondents' time if the research study was not robustly conceived or not completed. The main risk to non-completion of this study was the constraint on the time available to the author for completing the research and writing up this thesis whilst working full time.

4.7 Sampling strategy

For the interview study, the population was all the students in Y5 in the academic year 2014-15. These were stratified by their results in the end of Y4 OSCE in April 2014 and a purposeful sample was emailed directly and invited for interview. The approach was what Kuper et al. (2008a) define as 'critical case sampling' i.e. inviting participants whose information is predicted (based on a theoretical model) to be particularly illuminating.

For the first survey study, the population was all the students in Y4 and Y5 in the academic year 2015-16. The sample was a cross sectional convenience sample of students who attended specific lectures in spring or autumn 2015.
For the second survey study, the population was all the graduands (i.e. those students who had taken the final assessments and been advised that they had passed overall) from academic year 2015-6 before they graduated, and all students in Y4 and Y5 in the academic year 2016-17. The sample was a cross sectional convenience sample of the total population who attended specific lectures in summer/autumn 2016, or who agreed to complete an on-line survey circulated to the population. All respondents answered the same questions on their propensity to reflect on their learning activities in general as in the Phase 1 survey. Graduands gave information on their engagement with feedback from their end of Y5 OSCE in May 2016. Y4 and Y5 gave information about their engagement with feedback from their end of Y3 or Y4 OSCEs taken in April 2016.

Sample size and structure is important for providing validity evidence and, for the quantitative studies, to be able to make reliable generalisations to the whole population. Further information about the size and characteristics of the total population and the sample size and characteristics for respondents in each study is included in the relevant chapters below.

4.8 Reliability and Validity

It is important that evidence is provided to show that the research methods chosen were implemented with rigour (by following certain quality criteria) in order to demonstrate the credibility of the findings presented. In quantitative research

methods this is partly achieved by demonstrating evidence of reliability and validity. However, these concepts are not relevant to qualitative research methods, which by their nature generate data which are subjective, interpretive and context bound. Ringsted et al. (2011, p699) state that the notion of 'reliability' of data is replaced with a concept of 'trustworthiness' (via good analysis) and 'validity' with 'authenticity' (via strong data). Other elements of quality which they note include the adequacy of sampling and the 'saturation' of data, various forms of triangulation (examining the relationship of fit of data collected in various ways from various sources), member checking, multiple coding and the presence of an audit trail (record of decisions and work done).

4.8.1 The trustworthiness and authenticity of interview findings

I have adapted the guidelines on how to critically appraise qualitative research suggested by Kuper et al. (2008b) to make the case for the trustworthiness and authenticity of the findings from the interview study.

- **Was the sample used appropriate to the research question?**

The participant sampling strategy is noted above and the characteristics of the respondents are discussed in Chapter 5. Kuper et al. (2008a) raise the issue of sample sizes and note that qualitative studies do not usually have pre-determined sample sizes. Instead, sampling stops when the researcher feels that they have a thorough understanding of the phenomenon being studied. This is known as 'data saturation' and was developed in the context of Glaser and Strauss's (1967) Grounded Theory approach whereby researchers were believed to be able to identify the point of theoretical saturation, or when additional data do not add anything new to a developing theory. The term is now often applied to 'data' and 'thematic' saturation in studies which do not use the Grounded Theory approach. O'Riley and Parker (2012) and Varpio at al. (2017) are critical of the term being routinely used as a metric for quality without including a discussion of what this means in an individual study. I do not claim to have reached 'data saturation' in this study. Although there was agreement between what participants within and between performance categories reported that they thought and did, they still reported sufficiently different views that I would have preferred to interview more

students. However, despite sending out a reminder email, no further volunteers responded.

- **Were the data collected appropriately?**

A description of the context of the research setting has been provided in Chapter 1 so that readers know about the clinical examination and how the feedback to students was generated. The methods used to collect the data are provided above and in Chapter 5. The semi-structured interview schedule is shown as Appendix 2. Kvale describes an interview as a professional conversation that has a purpose and structure. He notes that the research interview is not a conversation between equal partners because the researcher defines and controls the situation (1996, p6). I am a senior officer of the University who is knowledgeable about assessment principles and processes, and obviously older than the students. This may have affected what the students chose to tell me. However, every participant appeared to answer the questions spontaneously, and there appeared to be internal cohesion in each student's interview narrative and no apparent attempt to appear to be conforming to an expected norm or self-aggrandisement. The methods used for interviewing and transcribing have been described above. However, as is usual, primary evidence for this (e.g. audio file or transcripts) are not available for external scrutiny.

- **Were the data analysed appropriately?**

I have described above a commonly used approach to qualitative data analysis which I followed. Studies which involve more than one researcher describe the process by which they ask a second researcher to undertake and verify the coding framework or themes (for example, Harrison et al. 2014, p233). In retrospect I could have asked another person to read and code some of the transcripts to check if they coded them in a similar way to me but I had not included this approach in the initial research design so did not have the participants' permission to share their primary data in this way.

Creswell (2014, p201) advocates that qualitative researchers should present their initial findings to the participants to confirm with them that the study reflects their views and experiences. However, pragmatically this was not possible because the

students who were interviewed graduated from the University before the chapter was fully written up. With regard to this study, this step was not as important to undertake as for emancipatory or action research studies where the results of the research may directly impact on the participants. Furthermore, Varpio et al. (2017, p46) make the point that for research underpinned by approaches such as constructivism which contend that data analysis is an interpretive process in which the researcher is actively part of constructing the findings and conclusions, no problems are created if the participants' and researcher's interpretations are different. Varpio et al. (2017, p47) also cite reasons why member checking may not be appropriate.

Silverman (2000, p11) raises the problem of 'anecdotalism' with regard to the credibility of evidence from qualitative research, which he explains as over reliance on 'a few telling examples of some apparent phenomenon without any attempt to analyse less clear or contradictory data'. However, I am not attempting to prove or disprove a hypothesis by selecting quotes which support a specific standpoint; I am exploring if a conceptual framework assembled from other models appears to be useful and if published findings from other contexts has resonance with the views of learners in this context.

- **Are the results of the study transferable?**

Quantitative studies normally include a discussion of the reliability of the findings i.e. the possibility of generating the same results if a very similar study were to be undertaken by different researchers, or if the results from a sample population could be extrapolated to apply to the whole of that population. Pope, Zeibland and Mays (2000, p114) note '… qualitative studies are not designed to be representative'. However Kuper et al. (2008a) use the concept of 'resonance' to describe the extent to which findings from this type of qualitative study have meaning for (resonate with) other situations and the confidence with respect to the applicability of a study to other settings. The results of this interview study could be compared with the results from interviews about another course where similar feedback is provided in order to explore similarities and differences and consider what factors may cause these. The findings from this study could also be used to inform the situation on similar courses where the introduction of individual

feedback is being considered, as OSCEs are commonly used methods of summative assessment in healthcare professional courses. The resulting conceptual framework which extends the usual Self-Regulated Leaning (SRL) cycle to include how effective SRL engage with feedback after assessment should be applicable to other learning situations.

- **Does the study adequately address potential ethical issues, including reflexivity?**

Ethical considerations are discussed in section 4.6 above. In addition to the usual ethical considerations, Kuper et al. (2008a) highlight the potential effects of power relationships between the researcher and participant. This is addressed in the section above. 'Reflexivity refers to the recognition of the influence a researcher brings to the research process' (Kuper et al. 2008b). With regard to addressing the possibility of bias due to the fact that the researcher is the main instrument in qualitative research (Walford 2001, p98), I reflected on my potential emotional involvement with the production of the feedback in order to attempt to avoid being judgemental about what the students reported to me regarding their attitudes and actions regarding the feedback and their studies in general, and to report the findings in a fair and balanced way. During the interviews, I recognised and internally acknowledged my feelings of frustration with the immature approach to their future professional practice signalled by one or two of the students' answers, and so I believe I was able to maintain a non-judgemental approach externally, and create rapport with each student to try to encourage them to give honest responses about what they thought or did.

- **Overall, is it clear what the researcher did?**

The reader should be able to follow the approach I took, understand the rationale for the decisions I made and come to a conclusion about the veracity of the findings and appropriateness of the consequent use to which they were put along with the recommendations I make in Chapters 5 and 9 as a result of interviewing the students.

4.8.2 The validity and reliability of survey instrument findings

Validity is concerned with the extent to which an instrument measures what it is intended to measure. Reliability is concerned with the ability of the instrument to measure this consistently, i.e. to produce similar results if the instrument tested the same population again. There are a range of aspects of validity which may be examined when considering the validity of an instrument. Messick (1994) and Kane (2013) argue that these elements should be assessed together as part of the process for building evidence for the appropriateness of the method used, in a unified approach to defining validity. Streiner and Kottner (2014, p1976) agree that validity is a unitary concept. 'It is the degree to which all of the accumulated evidence supports the interpretation of the test scores for the intended purposes'. A survey instrument per se cannot be held to be valid, it is the scores generated via the survey and the uses to which the scores are put which can be assessed for their validity (Messick, 1994).

However, DeVellis points out that Messick's six type classification of validity has not been widely adopted and argues that there are three essential types of validity (2017, p83). These are, validity inferred from the way in which the scale was initially constructed (content validity); a scale's relationship to measures of other constructs (construct validity); and the ability of the scale to predict specific events (criterion-related). Criterion related validity refers to an empirical association of new scale with 'some criterion or putative 'gold standard' (DeVellis, 2017, p92).

Validity and Reliability of the phase 1 and 2 questionnaire items

Clark and Watson (1995, p310) emphasise the importance of construct validity and note there are at least three steps to investigating the construct validity of an instrument; firstly to articulate a set of theoretical concepts and their interrelations (the nomological net), secondly to develop ways to measure the hypothetical constructs proposed by the theory and thirdly, empirically testing the hypothesised relations among constructs and their observable manifestations.

For my two studies, there was a strong nomological net in terms of the underpinning theoretical framework used, but as the usual processes for scale development were not followed for the EWF items, the 'measurement of the

hypothetical constructs' was weak. I took a pragmatic approach to the development of the EWF items because of lack of time, and devised the items myself, so the sub-scales may not include all aspects of the concepts they intend to measure, and so may lack 'content' validity. Streiner et al. (2015, p350) advise that a panel of content experts should be asked to review the survey items and for example, rate them on a scale of relevance and to ensure that the questions included covered all aspects the experts expect.

I used factor analysis to empirically test the hypothesised relations among the constructs, or sub-scales and used the results to refine the sub-scales. Clark and Watson note (1995 p311) 'good scale construction is an iterative process involving several periods of item writing, followed in each case by conceptual and psychometric analysis' and advise to 'err on the side of over-inclusiveness'. In retrospect it would have been better to include more EWF items in the Phase 1 survey because subsequent factor analysis demonstrated that two questions did not perform well and so were deleted after Phase 1. Streiner and Kottner (2014, p1976) caution that 'Newly developed scores or instruments should not be validated using the data set with which the instrument was developed. This usually leads to overly optimistic results'. Thus it is not possible to fully validate the EWF scale from these two administrations.

Cronbach's alpha coefficient is used to report the internal consistency of a scale (i.e. the degree to which all the items of a sub-scale measure the same attribute or dimension). This ranges between 0 and 1, with a score of .70 or above generally considered as being satisfactory (de Vaus 2014, p184). The alpha coefficient is affected by the number of items on a sub-scale. The number of items on the EWF sub-scales ranged from three to six, and this low number may partly account for the lower scores reported.

When calculating Cronbach's alpha, most software packages offer the option to also calculate the alpha score with each item in the sub-scale removed so that it is possible to check if retaining an item reduces the internal consistency and consider if overall, it would be better to remove the item, and I took cognisance of this when assessing the factored EWF sub-scales.

One of the ways to explore the reliability of a survey is to administer it to the same group of respondents at a later date. This type of test of temporal stability (test-retest reliability) of the same participants could not be conducted for these two studies because of feasibility issues (the responses were anonymous, it would not have been possible to assemble the same convenience sample of participants again, and some of the students graduated from the University later in the year and so would not have been contactable).

Streiner and Kottner (2014, p1975) highlight that 'Reliability is not a fixed immutable property of the scale; it is an interaction of the scale, the circumstances under which it is given and the specific group being assessed'. Therefore it is important to report the results for reliability every time a survey is administered and I have used Cronbach's alpha to do this.

4.9 Summary

This chapter discusses the rationale for choosing a mixed methods approach for this thesis; an initial qualitative interview study to assess and develop the proposed theoretical framework, followed by two quantitative survey studies in order to assess the respondents' propensity to reflect on their learning experiences and to develop and test an instrument to assess the respondents' engagement with their individual feedback comments. The chapter considers the ethical considerations relating to these research methods. The sampling strategies for the three studies are reviewed. Issues relating to providing evidence for the trustworthiness and authenticity of the qualitative method and the reliability and validity of the quantitative method are addressed. Consideration and careful implementation of all of these aspects helps to ensure the robustness and credibility of the findings presented from the empirical studies.

CHAPTER 5: STUDY 1 - INTERVIEWS WITH STUDENTS

5.1 Introduction

This is the first empirical study in this thesis and is framed in an interpretivist or constructivist research paradigm (Lingard and Kennedy, 2010, p324). Constructivism is predicated on the idea that all knowledge and shared reality is constructed by human interactions. In keeping with this epistemology, data were collected via semi-structured interviews with individual students to gain a nuanced understanding of their lived experiences, views and actions regarding this feedback. I acknowledge the impact of my own lived experiences and pre-conceived ideas in influencing the questions I asked, how I interacted with the participants, how they perceived me and my role and how I subsequently analysed the data; all of which have influenced the results presented below.

Walford states that 'all research is researching yourself' (2001, p98) in support of his assertion that all research has a subjective element, and especially so in qualitative research where the researcher is the main research instrument. In the following sections, a first person narrative approach has been used to emphasise the researcher's reflexive role in collecting and analysing this type of qualitative data (Thomas, 2013, p272). This chapter should be read in conjunction with Chapters 3 Conceptual framework and 4 Methodology which explain why this method of research was used and the rationale for the questions asked in this study.

5.1.1 Aims of this study

This study was conducted with two main, interconnected aims. Firstly; to provide scoping information about what students did (if anything) with the individual feedback comments provided after their end of year summative clinical examination and if they valued or preferred this type of feedback to the other types provided. Some of the published studies I reviewed reported on barriers to engagement with feedback, and an outcome nested in this first aim was to ascertain if the students felt these barriers were relevant to them. The research questions listed below relate to this first aim. Secondly; to assess if the outline

self-regulated learning (SRL) conceptual framework which I had previously developed (Figure 3.3 above) was useful for this context and if so, to further develop it to illuminate what effective engagement with feedback encompasses in this learning context. The outcome of this is shown as Figures 5.2 and 5.3 below.

5.1.2 Research Questions

1. How do students in each of 3 levels of performance engage with personalised feedback from a clinical examination?
2. Do students feel there are any barriers to engaging with the feedback?
3. Are any of the other types of feedback provided after the clinical examination felt to be more useful by students in each performance category and if so, why?
4. Do students feel that they need a structured system to support their engagement with the feedback?

5.2 Research Design

Semi-structured interview schedule

The process for the development of the interview questions has been described in Chapter 4. The interview schedule is shown as Appendix 2.

Participant recruitment

As noted in Chapter 4, the sampling strategy for participants in this study was a purposeful approach. The results of April 2014 first sit Y4 OSCE were used to stratify relevant students into three groups (Total population=380, female=63%)

1. Top performers (12 questions passed and average score ≥86%) N=15, female=60%
2. Borderline pass (8 or 9 questions passed and average score of 50-60%) N=16, female=63%
3. First sit fails without accepted extenuating circumstances and who subsequently passed the resit examination (≤7 stations passed and/or average score less than 50%) N=14, female=36%

The one Y4 student who was required to withdraw from the course after failing the

main and resit examinations was not contacted. Although s/he could have potentially provided interesting information about unsuccessful engagement with feedback, it would have been difficult to contact them in practice and importantly, it may not have been possible to maintain anonymity for the student in this situation.

In autumn 2014, all 45 students from these three groups were invited to attend for a one hour long, individual, semi-structured interview at a mutually convenient date and time in a private room in the Medical School. Students were not offered any inducements to volunteer to be interviewed.

5.3 Participants

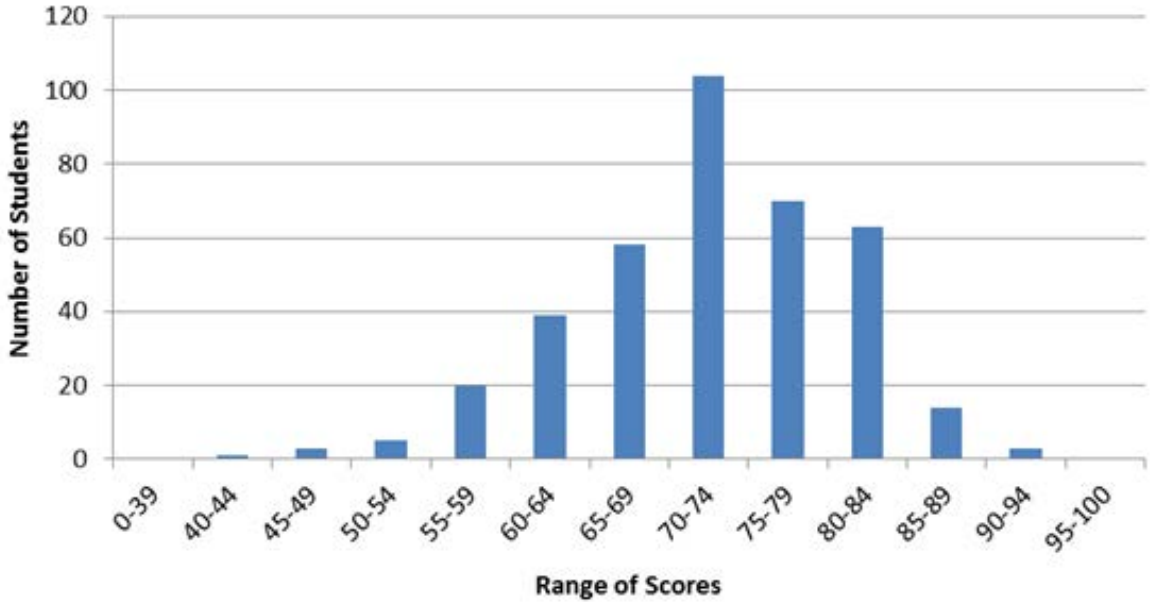
11 students agreed to participate and were interviewed (out of the potential group of 45 students from a cohort of 380). This group comprised one male and three female top performers, four female borderline performers and one male and two female first sit fail students. Table 5.1 below shows the participants' performance in the Y4 OSCE 2014. Figure 5.1 below that shows the OSCE scores for the whole cohort. The process for data analysis of the audio recordings of the interviews has been described in Chapter 4.

Table 5.1: Study participants average Y4 OSCE percentage score and number of stations passed (out of 12)

Performance level	Average Main sit % Score over 12 stations	Number of stations passed	Average Resit % Score over 12 stations	Resit No Number of stations passed
Top	93	12		
	89	12		
	87	12		
	86	12		
Borderline	60	9		
	60	9		
	60	8		
	59	8		
Resit	59	7	72	9
	54	7	71	10
	56	5	65	9

Note: To pass the OSCE, students have to achieve at least 50% overall and pass at least 8 out of 12 stations

Figure 5.1: Y4 Cohort OSCE average percentage scores 2014



Before each interview started, the student was given the same information sheet that was circulated with the email requesting volunteers to participate in the study and asked to sign this to document their consent to the interview and their permission for the interviews to be digitally audio-recorded and subsequently transcribed for analysis. Prior to being interviewed, all the participants had agreed that their individual feedback (their marks and the written comments per question they sat) could be printed out and brought to the interview as a basis for the discussions. The feedback that was circulated to all students was also printed out so that students could refer to it and refresh their memories of what it was and what they thought about it.

As detailed in Chapter 4, to maintain anonymity student responses were anonymised by being given a code comprising of the order they were interviewed in, followed by their performance level (T= top; B= borderline; R= resit) and their gender. E.g. 2TF= second student interviewed, top performer, female.
For information, the order in which students were interviewed is shown in Table 5.2 below.

Table 5.2: The order in which students were interviewed

1	2	3	4	5	6	7	8	9	10	11
BF	TF	TF	TM	TF	RF	BF	BF	BF	RF	RM

Note: Performance: T= top, B= borderline, R = resit; F=female, M=male

5.4 Analyses

As noted in Chapter 4, a thematic coding analysis as described by Robson (2011, p476) was conducted. This resulted in the following initial codes:

- Comments on FB by students positive (accepted, actionable, said did act on, matched own view, even when examiner made negative comment)
- Comments on FB by students negative (not accepted, not actionable/understandable, not match own view)
- Digression in interview (to exam process, standard setting, mark sheets, examiner training).
- Discussed written FB with others
- Emotional response to FB (ref decision to look at it or not, when reading it or after reading it)
- Goal setting; including any comments on approaches to study/revision and marks that aimed to achieve
- Implementation of a FB support system
- Legibility of FB comments
- Marks; including as a short cut to confirm self-assessment, satisfaction with marks achieved
- Other types of FB than the individual written comments
- OSCE format – performance and nerves
- Problem examiners or questions
- Reflection on initial thoughts about FB comments
- Resit Exams (all comments relating to)
- Self-assessment of performance directly after exam on own or after discussion with other students.

These codes were conflated into the following themes which informed the development of the initial conceptual framework and enabled the research questions to be answered.

- Self-assessment/Reflection before FB released (including how student did this, by self or via interaction with others, and why - to improve clinical

practice or estimate pass or fail).

- Action taken to improve before FB released
- Comparison of self-assessment and examiner assessment in feedback comments, reflection on/reconciliation of any differences (including how student did this, by self or via interaction with others).
- Action taken to improve after reflecting on FB comments
- Barriers to engaging with comments (legibility, examiners, OSCE questions, emotions)
- Preference for other types of feedback
- Structure to support engagement with feedback

Where appropriate, in the analysis below, I have indicated where there was consensus between students of the same performance level.

5.4.1 How do students in each of 3 levels of performance engage with personalised feedback from a clinical examination?

Following the cycle of events depicted in the final version of the conceptual framework (shown in Chapter 5, Figure 5.3), starting with the step 'Student reflects on and self-assesses performance in the clinical exam before receiving the examiner feedback. Student may also act to improve self-assessed deficiencies', it was apparent from the interviews with the four top performing students that they followed these actions. In response to my question 'Did you look up anything after the OSCE exam you thought you got wrong before receiving your marks or the written feedback?' *4TM said 'Yeh definitely. I think I do that with almost every exam. I try to remember questions that I got stuck on and those are the ones you remember more easily anyway and I always go and look up the answer to those. I often think there are some questions that you wouldn't be able to find but I do try. I am trying to think of an example...one of my stations was peak flow so I went to look up the normal ranges…. I looked it up even though I don't think it's useful to know because you always look at the normagram, but I looked at it anyway to see if I was right or wrong'.*

Borderline and resit students indicated that that they did not tend to act to improve any perceived deficits after the OSCE but before accessing feedback. 8BF said '*I didn't look at a text book because I thought it's done and there's nothing I can do about it*'. 11RM looked up one thing because he felt embarrassed at his lack of knowledge during the station. '*Not with a view to anything serious, but on Wikipedia, I looked up a type of psoriasis that I had never heard of but the station was about and I thought OK, that's what it's about'.*

When reflecting on their performance and self-assessing strengths and weaknesses, top performing students did not tend to discuss their performance with other students. 3TF said *'I don't usually speak to people who sat the same questions as me because I don't really like the hysteria. I come out of the exam and the things I look up are the things I thought I should have known that and I didn't. And it's for my own learning in the future that I'm going to go and look that up and try not to look stuff up like, did I get this right, did I get that right?.'*

The Resit and Borderline category students discussed questions with other students to gauge if they had passed or failed rather than to assess perceived strengths and weaknesses. 6RF said '*Yes. I always do that though. I literally went through it with my friends and they hated me for it, going through questions, what we could remember trying to figure out what we got right, what we got wrong and you kind of get an overall feeling if you've got a pass or a fail.'*

It has been noted in the literature how difficult it is for learners to accurately self-assess their strengths and weaknesses (e.g. McConnell et al. 2012; White, Ross and Gruppen 2009; Srinivasan et al. 2007). In keeping with the results of other studies, the top students tended to under estimate their performance compared to the feedback subsequently received. 2TF said *'After the OSCE I had at least 2 or 3 stations I managed to convince myself that I had failed because I am a perfectionist, so for me if I can't answer a question or I'm not happy with the way that I answered a question then I tell myself that I have failed that station which I know is ridiculous and is definitely the worst case scenario but for some reason I would rather feel like that I'm erring on the side of caution*'.

The poorer performing students tended to overestimate their performance. 11RM said '*I didn't specifically try to work out how many marks as I think it's difficult to know how people are going to grade you, but I did have a good feeling about whether I had passed a station or not'.* However, he also reported in the interview that that he was surprised when he learned he had failed the OSCE.

Moving on to the stage in the conceptual framework after feedback has been released and potentially accessed; 'Student accesses and critically reflects on the feedback which may contradict, confirm, add to or modify their existing domain or metacognitive knowledge, beliefs about self or learning strategies. Student reconciles internal feedback (self-assessment) and external feedback and selectively acts on the feedback comments before or during next formal learning cycle'. The four top performing students accessed the feedback as soon as possible after it was released. They critically reflected on the comments, both on what they did well and could improve on and considered if they agreed with them or not.

5TF *'I thought that some of them were fair, others of them I thought no way, that didn't reflect me. And others I felt 'Had the person who examined me actually listened to what I said?' because they've written something completely different to what I've said – but I know a lot of people feel like that so I suppose in some ways you then try to take the best bits you agree with from it but then other bits you don't agree with you say well I know factually that that didn't happen so I'm not going to use it'.*

Sometimes top students ignored comments under 'what could you improve', but had a logical rationale for this. For example 3TF said *'The other station that I didn't do as well on was the HIV station which I think was my lowest one. It was one of those things where I'm obviously quite a talker and not everyone will like my style of how I go about things and it was just one of those things where I got the sense that the examiner didn't like how I went about doing things. So it was one of those things where you can't help yourself doing it, but the examiner was thinking if I'd have done it, I wouldn't have done it that way, and I just think I did it in a way that he wasn't a fan of'.*

Top students acted on other critical comments, even where these did not fit with their own perceptions of their performance. E.g. 5TF said '*And they've told me I used jargon again and I said to my other half I genuinely don't know what that is because I can't remember saying anything but if multiple people are saying it then I must do it so I need to be aware of it.'*

The top performers used the feedback as feed forward for Y5 and clinical practice. 2TF said '*And it also made me think, the fact that there were two that I'd forgotten to ask about family history, that is something that I need to drill into myself for the next time 'cos it's not acceptable to be forgetting that in real life'*

Borderline students were less homogeneous in their response to the feedback but did provide examples of accessing and acting on the feedback. For example, 8BF did not act to remediate her knowledge deficiency before accessing the comments, but reading the feedback comments triggered action and the student thinks this knowledge will remain with her for the longer term.

Me [Read the 'What could you improve' feedback comment out loud.] What do you think of that?

8BF *Yes that's fair. Obviously at that point I didn't link it together. I went back to my books and realised oh yes, how could I forget about this (the link with Ankylosing spondylitis - AS)*

Me So when you read the feedback, is that when you went back to your book?

8BF *Yes and its stuck with me now that with Uveitis think about AS. When I don't know the answer on a station and someone tells me then I make the link and I won't forget it again.*

Borderline students also gave examples of reflecting on the feedback comments, although not necessarily acting on them. For example, 7BF said *'When I first read the comments on the stations that I failed, I felt very irritated about it all because I couldn't work out why I'd failed as the good feedback made me look as if I would be a safe doctor so I was a bit angry towards it. But as time's gone on, I'm maybe more able to understand that perhaps I didn't come across as confident, I did*

hesitate a lot with for example the investigations, I can see that the stations didn't go as well as some of the other ones which I didn't take into account at first because I was angry'.

One of the borderline students, 9BF, reported that she had not accessed all the feedback because the first comment she read under 'What did you do well' described her as friendly to the role player and she dismissed this as being condescending.

Not surprisingly, the resit students all accessed the comments. Two of the resit students did not appear to understand that feedback could be used to enhance future learning and clinical practice; they only saw the value in helping to pass an assessment hurdle.

10RF did not appear to reflect on the comments at all. She discounted both negative and positive feedback where she felt her mark had been reasonably high because '*Given that question again, I would perform the same'*. Three examiners had written feedback regarding her interactions with (role played) patients (e.g. 'try to build more rapport with the patient', and 'argumentative consultation style'). When we discussed what she thought of this she simply said '*Well, I think I am empathetic*'.

When we discussed his feedback from each station, 11RM appeared to only see the value of improving his knowledge in case he was tested on it again in an examination, not for future clinical practice, for example; '*It was useful, I guess, to know what to do differently if I came upon a skin station again'.*

6RF was different to the other two fail students in her approach to engaging with the feedback. She was reflective and had analysed the comments and showed me the notes of the action plan she had made after initially reading her feedback. She said *'You've got 12 pieces of paper each saying different things but what it boiled down to was problems with my knowledge, history taking and being organised, structured…. and I can do that, it's just the nerves. I think I just blurt out anything that comes in to my head. So it's all about controlling nerves and*

having the knowledge'. As a further example of her reflexivity and action planning capability, she had also self-referred for professional advice on building her confidence for performing in future clinical examinations.

With regard to the feedback from the resit examination, when asked if she had done anything different with the resit feedback comments compared to the first sit ones, 10RF replied *'Yes because after first sit I read it and took it into account whereas after resit I thought I don't need to read it because I passed'.*
11RM said *'When you get the email saying the marks results are up then you think oh my gosh and you find out you've passed and you think Ok I don't need to look at it'.*
However, 6RF was again very different in her approach, *'I went through it the same as first sit. I didn't do anything different and I still thought it was really useful. If you look at my marks from resit, I still failed a few stations'.*

5.4.2 Do students feel there are any barriers to engaging with the feedback?

A previous review of the literature suggested that there were a number of factors which might act as 'perceived barriers' to students' engagement with feedback. During the interviews, these potential barriers were explored to ascertain if students demonstrated tendencies towards 'extrinsic motivation to learn' (White 2007, p281), which might have biased their views about the veracity of the feedback comments, and provided a self-justified rationale for not acting on the feedback.

These 'barriers' include legibility of handwritten comments (Agiusa and Wilkinson, 2014). In the study presented in this thesis, at least 130 samples of handwritten feedback comments were discussed and the students felt that most comments were legible. There were only 13 negative comments from the students about legibility because of handwriting. 3TF noted *'Yeh some of it was Ok, some of it took a while to decipher, but the majority of them I could read. Yeh they did a better job than I'd have done.'*

Other studies have reported that the credibility of the feedback provider can be an issue in accepting feedback (e.g. Bing-You et al. 1997, Watling and Lingard,

2012). However, students reported few perceived problems with examiners which might have then have impacted on their willingness to engage with comments from them. For example, 7BF said *'I don't think anyone was overly stern'.'*

Patel et al. (2015) noted that failing students sometimes blamed the 'unfair' question topics in an OSCE for their failure, rather than their lack of expected knowledge or skills. However, in this study, the students felt that the majority of the questions were reasonable topics to be examined on. Feedback from 24 OSCE questions were discussed with this sample of students and students only expressed negative views on specific tasks or questions within three questions in which they were expected to know; the normal values for an adult peak flow reading; the diagnosis of sick sinus syndrome from the symptoms reported; and the colour of oxygen cylinders.

Various studies have highlighted the influence of emotions in being ready to accept feedback (e.g. Sargeant et al. 2008; Sandars and Cleary 2011). In response to the question 'When you read the email that said the individual written feedback comments from the OSCE were available, what did you think/how did you feel?' Top students had positive emotions about reading the feedback. 2TF said *'I was waiting for it to come out!'* and 4TM said *'I was interested to see what they had said about me'.*

Borderline students reported they were worried about what the examiners had written about them, but when they did read the comments they were surprised. 1BF said *'I thought it was good. I didn't look at it for a while because I knew I hadn't failed but it affected my confidence anyway and I didn't want to read it because it would make me feel really crap. But I did look at it and I thought hold on a second…'*

8BF *'I was quite nervous thinking about it. I didn't want to see the ones for the failed ones! I was away but I went straight to look at it because I thought well the exam's done already so you might as well look at it. I think it was quite good because they not only comment on your negative things, they also include the*

positive things, which is really good because it didn't make me feel like I'm a terrible student'.

Interestingly, the resit students were more positive about reading the feedback compared to the borderline pass students. They did not express angst at the thought of reading the negative comments. This may have been because they welcomed the feedback to guide their revision.
11RM said *'I was like yes, I want to see that'.*
10RF said '*I thought it would be a good thing to see any justification behind the marks that were given. I was looking forward to reading what was said regarding each station'.'*
6RF said *'So glad because clearly I had failed my OSCE so I wanted individual feedback'.*

5.4.3 Are any of the other types of feedback provided after the clinical examination felt to be more useful by students in each performance category and if so, why?

Ten of the 11 students preferred the individual comments. For example, 6RF said '*Yes on the whole I would say that the individual feedback, if it was legible, was incredibly useful and it was good to know what I got right or wrong'.*

9BF strongly preferred the generic written comments (although she had not read all of her individual comments previously, we did review them all during the interview). *'The generic feedback is way more detailed, which is probably more useful even though it's not personal. So if you compare the HIV one, there are actual specific points that I personally could go away and think about, like don't use medical jargon, don't forget to check the partner's gender, and so on, so it's way more specific than the individual feedback...'*

Students also compared their individual comments with the generic ones to assess their own strengths and weaknesses against the cohort. 2TF said that she did this '*to get the maximum benefit from the feedback*'.

The average scores per question were mainly used for benchmarking. 7BF said '*There's some benefit to see that these scores are achievable, you can see it's not all because of the examiners. It gives you a sense of responsibility when you are trying to offload blame for doing badly'.*

9BF said *'I suppose it helps you to know what the general standard is. But if you're not at the general standard it doesn't make you feel inspired.'*

Where they looked at it, all students used the histogram of overall scores 'competitively' to find their decile ranking (used as part of the national process to allocate junior doctors to their first training post); they were all clear that this did not help to improve their learning. 4TM said *'I think the histogram is useful but only from a competitive point of view. I don't think it advances my own learning in any way'.*
7BF said *'The histogram depresses me!'*

5.4.4 Do students feel that they need a structured system to support their engagement with the feedback?

There was limited support from the 11 respondents for the idea of introducing a structured support system, and that was mainly from the lower performing students.
5TF said *'No, I think it should be enough just to provide the feedback...you don't need to be spoon fed a way to use that information.'*
10RF agreed '*If someone has written there's a gap in your knowledge you should know by yourself how to fill that gap, especially at this stage in Y4 or Y5. You don't know the gaps in your knowledge until they are pointed out, but once it is you should know how to go about it, it's not a problem.*'

This lack of support for a structure to be implemented by the course organisers does not appear to be because students have evolved alternative, peer or other support structures. The lack of reported peer support may be due to the competitive nature of medical students and them not wishing to reveal their weaknesses. 8BF said *'I felt like I didn't do as well as most people so I was like,*

not embarrassed but I didn't want to show people and have them think that I'm not at the same level as them. No-one discussed it, not my housemates even.'

Only student 7BF said that she discussed her feedback with peers in order to be able to understand it better: '*I like discussing things with people in an open discussion so your feedback and mine with someone else who could help you a bit more about why the examiner wrote that.'*

All students on the course are allocated a Personal Mentor (PM) and some of the Borderline and Resit students could see benefit in discussing the feedback with their PM, although they could have used their initiative and contacted their PMs to ask if they would be willing to discuss the feedback with them.

Student 11RM said *'Not really because the things examiners say are so variable it would be hard to design some generic support to help people to understand it, really. It might be useful to meet my PM tutor to discuss what he thought about it.'*

Student 7BF agreed '*Personal mentor system would be good because it gives people chance to vent a little bit and maybe be able to explain a bit more about things like being particular about holding the otoscope a specific way'*.

8BF said *'I think discussing with your PM would be a good idea. Not so much for me as I'm in final year already but in earlier years, yes because if you had problems and you discussed them with your PM they would try to help you.'*

The initial outline conceptual framework (Figure 3.3 shown above) was further developed based on key points from the interviews (Figure 5.2), and the final version of the conceptual framework is shown as Figure 5.3 below.

5.4.5 Conceptual framework of processes in SRL incorporating engagement with feedback

Figure 5.2: Development of the outline conceptual framework of processes in self-regulated learning incorporating engagement with feedback contextualised to Y4 MB ChB at Birmingham

Pink box = process internal to student
Blue box = process external to student
Green box = information from interviews with 11 students stratified by performance level
Dotted line box = area of interest in this study

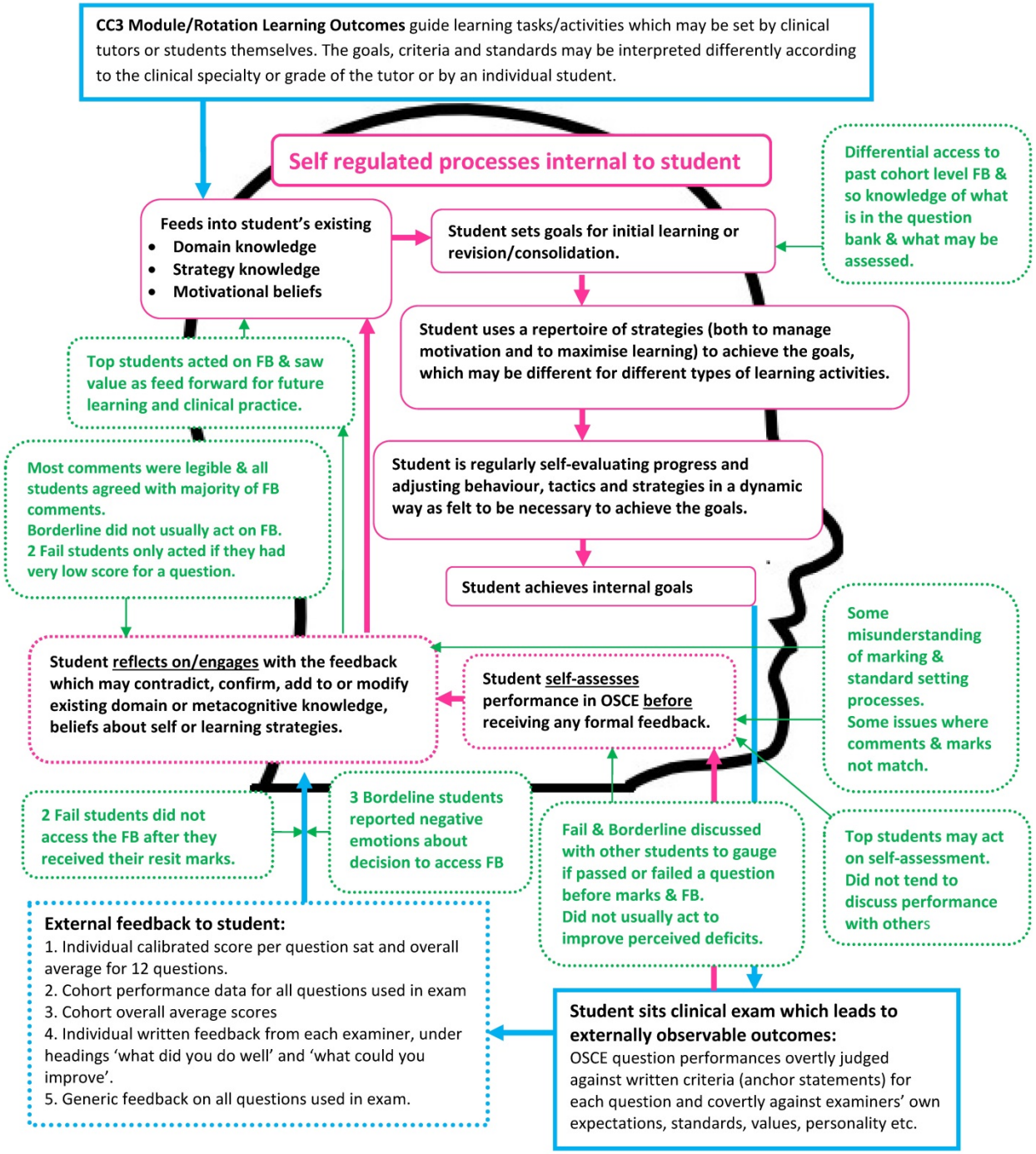
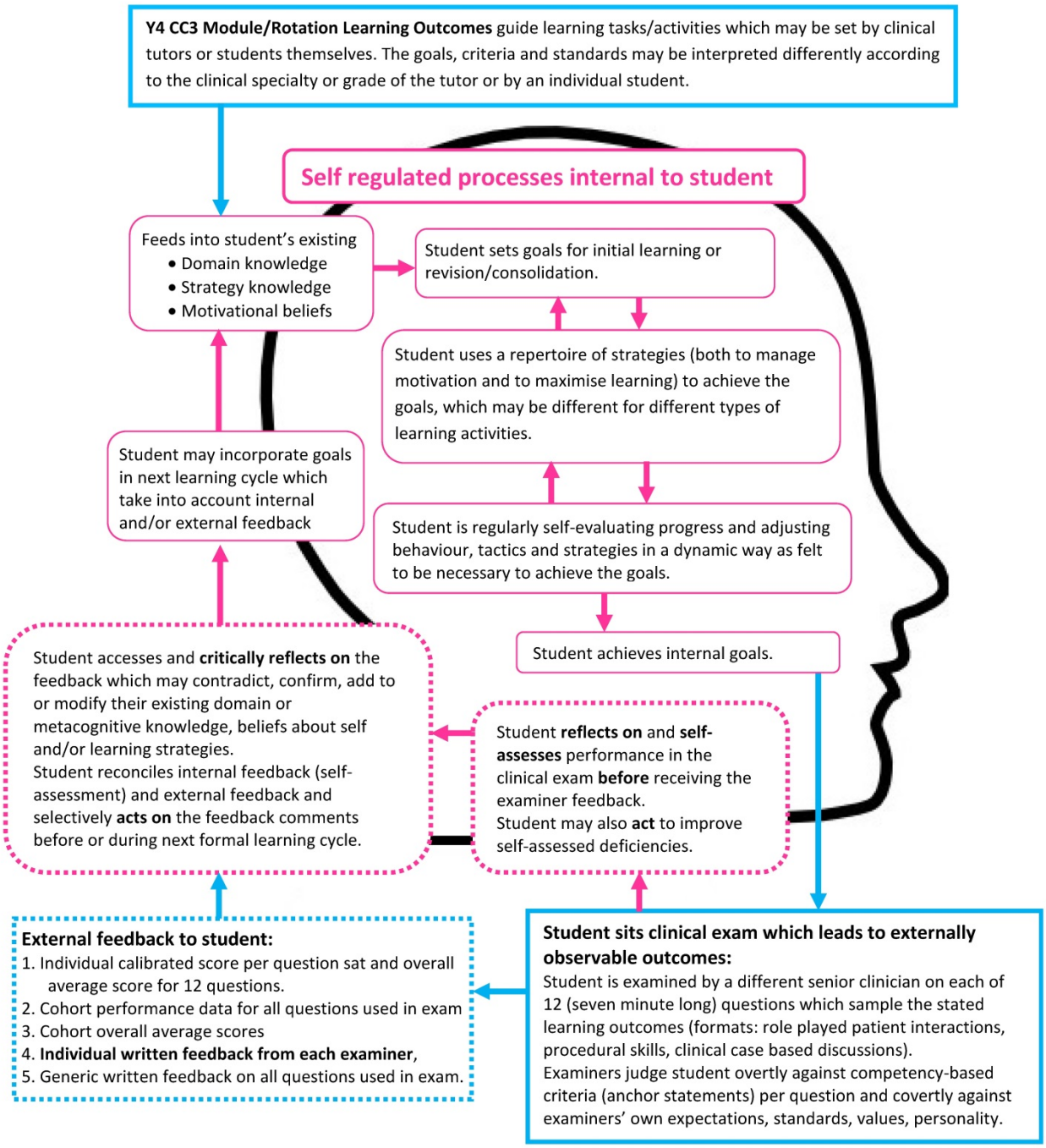


Figure 5.3: Conceptual framework of processes in self-regulated learning incorporating engagement with feedback contextualised to Y4 MB ChB at Birmingham

Pink box = process internal to student
Blue box = process external to student
Dotted line box = area of interest in this study



5.5 Discussion

I intended to be systematic in asking all the students the same thing, but I was flexible in allowing students to ask questions in return (e.g. about the examination processes) so that the interview flowed more naturally, and this meant that I did focus on slightly different things depending on the students' responses to my questions and the questions they asked me. The interviews focussed on exploring the steps in the outline effective self-regulated learning framework which related to self-assessment and action to improve perceived deficits in knowledge or skills *before* receiving any feedback; accessing the feedback and critically reflecting on this to reconcile any differences in the initial, self-assessed internal feedback with the external feedback comments; and subsequent action to improve self-agreed deficits in knowledge or skills. This exploration also included discussing factors that might encourage or discourage students from engaging with the feedback comments. The students' responses were analysed according to these steps and the actions of the top performing students were contrasted with those of the borderline pass and resit students.

Robson (2011, p281) notes '...just as you are hoping to get something out of the interview, it is not unreasonable for the interviewee to get something out of you', and this was true in that discussions with respondents often led to them asking questions about exams processes such as how the pass mark per question is set. It was clear that some students misunderstood aspects of the marking and standard setting processes. As previously explained in Chapter 1, each sub question or task within a question may be weighted differently and examiners do not know what marks they are awarding. After the OSCE, but before the release of marks or feedback, the Borderline and Fail performance level students tended to self-assess whether they had passed or failed a station, rather than focussing on their specific knowledge or skills. If a student did this by self-assessing that they answered half of the questions correctly, because of the differential weightings and standard set calibrated pass mark, contrary to their expectations, they may not have passed that question overall.

For example 11RM *'…after you have explained how different questions are weighted differently and now I understand there is much more rigidity in the marks*

and examiners don't know the marks at all, so they don't know they are giving you a 49, which I felt hard done by before as I thought they had purposefully failed me'.

Three of the four Borderline level students reported that they had read the feedback. However, these students did not often report that they had reflected or acted on the feedback to remediate deficiencies in knowledge or skills. During the discussions they said that the majority of the feedback was useful and actionable. However, they still viewed the feedback in the context of how they performed during the examination, not as actionable information which could be used to improve future professional practice. Some of the feedback comments were about the oral nature of the clinical examination and the students' 'live performance' in the OSCE (e.g. about their nervousness, or hesitation, or appearing confident) which reinforced the students' notion that the feedback was only about that moment in time, not potential 'feedforward' to improve clinical practice.

Student 6RF acted very differently to the other two resit students. Although she self-assessed by discussing her performance with her peers and comparing herself with them, as the poorer performing students tended to, she did reflect deeply on the feedback provided after the main sit OSCE and thought about deficits in her skills and knowledge at a metacognitive level as well as the specific points relating to individual questions. Unlike the other two resit students, she went through the same process with the feedback from the resit OSCE.

With regard to the second research question, all students preferred the written feedback comments, with all but one student preferring the individual comments over the generic ones. They were clear that the generic numerical information did not enhance their understanding of their knowledge and skills but were useful for benchmarking individual performance against the cohort's achievements. This preference for specific comments shows that the borderline students were interested in knowing what the examiners thought of their knowledge and skills, even if they did not act on this. The borderline student who preferred the generic comments said it was because she believed they were more detailed and so more useful.

A previous survey of the relevant literature implied that the provision of a dialogic structure to support the giving and receiving of feedback leads to improved understanding of the feedback and potentially, action to improve (e.g. Nicol and Macfarlane-Dick, 2006 and Boud, 2015). Contrary to my initial expectations, there was very limited support from the students for the provision of a support structure to facilitate engagement with feedback. This lack of enthusiasm may be because these fifth year students have progressed through the course so far without one and so cannot envisage how such a structure might operate. It may also be because these are senior students who are embedded in a strongly bounded community of practice (Lave and Wenger, 1991) and so generally understood the comments, even if they did not act on them.

Strengths of this study

This study includes responses from students from three different levels of performance, including those who initially failed the examination. After this study had been conducted, Harrison et al. (2014) published their findings from a similar qualitative study and reported that they were unable to recruit any students who had failed the main sit and only one student who was a borderline pass.

I felt that the students were all candid with me, perhaps because I am not a clinician, and so not identified as being in their 'Community of Practice' (Lave and Wenger, 1991), nor seen as an authority figure in the Medical School hierarchy. The participants all gave permission for their twelve pieces of feedback to be printed, and brought to the interview and these formed a framework for discussion rather than eliciting their views on the feedback in general. Although there was a time gap of approximately four months between receiving the feedback and the interview, this process helped to refresh the student's memory of each question, the feedback comments and their thoughts, feelings and any actions they undertook to improve skills or knowledge. The time gap also meant that the students were back on clinical placements so had access to clinical facilities and patients if they had decided to act to improve specific skills. This enhanced the validity of the conclusions drawn from the interviews.

Limitations of this study

Although sufficient information was gathered to be able to add detail to the SRL conceptual framework, I thought that 'data saturation' (O'Reilly and Parker, 2012) had not been reached for any of the performance categories, as I felt that each student was still giving new insights into how they thought or acted. Participants were self-selecting and the characteristics of students who volunteer to be interviewed may have introduced bias in the views presented from this sample. Participants from the majority 'good' performance category were purposefully not recruited for this scoping study which aimed to focus on contrasting the best with the worst performing students to more clearly highlight any differences in views. However, limiting potential participants in this way may have been a methodological flaw in terms of enabling a wider understanding of students' views about the feedback provided.

All four top level participants were graduate entry (with a degree in a science based subject, usually having achieved first class honours) and so were older and may have had a different approach to engagement with feedback compared to the top performers from the main course (usually direct from school) entrants.

The data set and findings are predicated on the author's choice of interview questions and the phrasing of these, which may not have been optimal to explore insights from students on how they engaged with the feedback. As Braun and Clarke (2008, p98) note 'The challenge for the novice researcher is to interact with research participants in such a way that they generate rich and complex insights.' On reflection, it might been useful to examine more fully and directly the participants' views on and approaches to exactly how they self-assessed their performance, and to have attempted to make more transparent the reflective processes by which they reconciled the positive and negative feedback comments with their own views, if they did do this.

Another limitation of this study may have been my interviewing technique. Topics raised by students were explored to gain a fuller understanding of their interests or concerns, and so the same data was not collected from all students via the questions on the interview schedule. Therefore some opportunities to collect data

that could be compared by performance category may have been missed. It might have been easier to keep to the interview schedule if the focus had been solely on the feedback comments and any discussion of the marks awarded had been avoided or curtailed.

5.6 Conclusion

The results of this scoping study enabled the further development of the outline conceptual framework. The processes for effective engagement with the individual feedback comments were drawn from what the top performing students in the Y4 OSCE reported they did before and after the feedback was released. This was contrasted with how the lower performing students described what they did. The resit students all engaged with the feedback, but generally were not as effective in terms of self-assessment before the feedback was released and subsequently reflecting on the comments after reading them. They did act to improve some of their deficiencies in knowledge or skills, but for two students this action was overtly framed with the intention to pass the resit OSCE, so they were selective and strategic in what they did, rather than acting on all the feedback to seek to improve for future clinical practice.

It was obvious during the interviews that the Borderline students were interested in the feedback comments. They reported that most of the feedback comments were understandable, useful and so potentially actionable, but they did not tend to act on them. More focussed questioning did not reveal any clear intrinsic reasons for this. In order not to appear to be judgemental, it was not elicited exactly why these students did not perceive the benefits of acting on the information provided, although it is possible to extrapolate this from some of the replies. The examination is generally only perceived as a summative hurdle to be successfully cleared and not as a potential learning event via the opportunity to audit and reflect on strengths and weaknesses. After I had conducted this study, Harrison et al. (2015) reported a similar individual interview based study with 17 participants. They noted analogous findings ‘The summative assessment culture, with a focus on avoiding failure, was a dominant and negative influence on the use of feedback’. Harrison et al. (2014, p229)

The final version of the conceptual framework for effective self-regulated engagement with feedback contextualised to Birmingham Y4 OSCE is shown as in Chapter 5, Figure 5.3. Although the interviews with all of the students were very interesting and greatly enhanced my understanding of the students' thoughts and actions regarding the feedback, there was insufficient data to be able to anticipate how students at each performance category in the wider cohort population might typically engage with the individual feedback comments. For example, with regard to the resit students and their approach to engagement with feedback as represented in the steps in the conceptual framework, the reflective nature of student 6RF was so different to the other two resit students. This highlighted to me that the views and actions of the students in the other two performance categories may also have been atypical and so I was not persuaded by this limited sample of data that there was a definable pattern to how students engaged with the feedback according to their performance level.

I viewed the performance level in the clinical exam as a potential proxy measure for the students' efficacy at self-regulated learning processes in general, although this study only considered their thoughts and actions on certain steps of this overall SRL conceptual framework. The propensity of a student to initially reflect on and self-assess their performance after the examination and before the feedback was available, and then to compare this to the external feedback provided from the examiners later, are two fundamental steps in the final conceptual framework. The top students were internally motivated to act to improve their clinical practice. The resit students were externally motivated to act to improve their performance in the resit examination. Therefore action, in terms of understanding the feedback and knowing what to do in these two different situations, should be possible for students at all levels of performance. The question is, how to make borderline students (and possibly students at other levels of performance) realise the importance of taking this step?

Many published reports indicate that feedback is more effective when provided within a dialogical framework (at least in most non-professional HE settings) because students may not be able to understand the feedback comments in terms of the norms of that discipline or know how to act on the feedback provided in

order to improve, and the discussion with tutors or peers ensured that students were cognitively engaging with their feedback (c.f. Price et al., 2011, Blair et al., 2014). Prior to conducting the interviews with students, I wondered if this could be the case in my context as well, and if an external system of structured support from faculty or peers would be beneficial in improving student engagement with feedback by providing an opportunity for dialogue and an element of policing, to ensure a measure of at least external compliance.

However, the students in this context did not report strong support for this idea. In the light of the discussions with the participants on this study, and further exploration of the literature, I started to consider if it would be more efficacious for the Medical School faculty to put efforts into inculcating students into understanding why they should, and how they can, actively make use of the feedback information provided after summative assessments and to develop a 'feedback culture' (Archer, 2010, Watling et al., 2013, Watling, 2014). Bing-You and Trowbridge noted 'Learner curricula could include training in how to recognize, receive, and respond to feedback at a metacognitive level.' (2009, p1330).

It would be more sustainable to provide students with guidance on how to effectively engage with feedback in terms of the self-regulated learning conceptual framework and to give students opportunities to practise and develop their understanding of how to do this. This would mirror the effort the Medical School faculty put into the students' development of other internalised professional behaviours. The assumption that students can be guided to be more effective self-regulated learners is supported by Nicol and Macfarlane Dick (2006, p205) and Sandars and Cleary (2011, p875).

5.7 Summary

This chapter describes how the first empirical study was conceived and conducted. The theoretical framing of the research and the researcher's role and voice in this type of data collection is made transparent. This study aimed to assess if the outline conceptual framework, which was an amalgamation of two published models of self-regulated learning, had resonance in this context and could be further developed with regard to how students reported they engaged

with feedback after the examination . Barriers to accessing or acting on the feedback which were reported to exist in other contexts did not appear to be relevant to this one. In the light of the information from students, the conceptual framework was further elaborated. One important step which was added after interviewing the students was to describe their approaches to self-assess their performance in the examination, and in some cases to act to improve based on this, *before* the external feedback was provided. The results of this study made me reconsider my initial ideas about how students could be supported to engage with the feedback.

CHAPTER 6: THE REFLECTION QUESTIONNAIRE

6.1 Introduction

As reported in Chapter 3, self-regulated learning (SRL) draws on a collection of metacognitive skills, habits or processes by which learners are motivated to set and achieve learning goals. The ability to reflect on learning events is an important component of effective self-regulated learning. Zimmerman posits that self-regulatory processes fall into three cyclical phases: forethought, performance or volitional control, and self-reflection processes (2005 p16). These can be summarised as the 'before', 'during' and 'after' phases of a learning event. It is assumed in this thesis that the framing of this cyclical process can be extended to the situation relating to clinical examinations. The 'before' phase being the self-regulated activities pertaining to learning or revising the chosen topics and reflecting on experiences of similar examinations; the 'during' phase being the active demonstration of knowledge and skills in the examination and the extended 'after' phase being when the student reflects on what and how they did or said during each question or station in the examination and self-assesses strengths and weaknesses against the expectations of the question. They may also decide to act to remediate perceived deficiencies in knowledge or skills. This reflective stage is revisited after accessing feedback from the examiners, when students may reassess what they initially believed about their performance, and may act to remedy deficits in knowledge and skills.

As previous research has shown that effective self-regulation (which includes the ability to reflect on learning activities and progress towards goals) is a distinguishing characteristic between high and low performing students in HE settings (Sundre and Kitsantas, 2004), it was hypothesised that students who performed well in the clinical examinations would have the skills to reflect on their learning in general, and assessing this would be a proxy for ability to also reflect on the activities in the clinical examinations. In order to explore this hypothesis, I decided to use two sub-scales from an existing survey, Kember et al.'s Reflection Questionnaire (2000), in order to assess students' propensity to reflect on their learning on this professional course of study. The aim of this chapter is to explore

if this previously validated instrument appears to be reliable and valid in this situation.

6.2 Existing Scales for measuring reflection skills

Various self-completion surveys that seek to assess the respondents' propensity for reflection have been reported in the literature (e.g. Mitchell, 1994; Kember et al., 2000; Sobral, 2000; Peltier et al., 2005; Aukes et al., 2007). Other approaches to measuring students' competence in reflection have also been reported, such as methods to assess students' written reflective accounts (e.g. Aronson et al., 2012; Pee et al., 2002). However, a self-completion survey was a feasible approach to this study's aim of measuring the reflection skills of a planned large number of students and comparing the results of this to the respondents' scores from the end of year summative clinical examinations in order to explore if there was a relationship between the two measures.

Initially I felt that Sobral's (2000) 'Reflection in Learning' questionnaire would be the most appropriate instrument to use for this study because some of the 14 items referred directly to facets of the self-regulated learning conceptual framework, Sobral developed this survey instrument with medical students (in Brazil) and reported satisfactory reliability and validity indices, and because its psychometric properties have also been comprehensively assessed and found to be acceptable in the context of trainee teachers (Kalk et al., 2014).

However, difficulties were encountered when attempting to choose and adapt an appropriate sub-set of these items for inclusion in this study, partly because of the style of English (they were originally written in Portuguese) but also because the items were devised as a one factor set. Kalk et al. (2014) propose a four factor model for the Reflection in Learning scale after analysis of the data from their administration of this survey (in which they assigned three items to a reflection sub-scale), but their item groupings and decision to remove certain items were not fully explained and did not appear to be intuitive.

Instead, I decided to use eight items (two factors) from Kember et al.'s (2000) 16 item 'Reflection Questionnaire' as the survey instrument for the following reasons;

the authors developed it with a robust sample (N=303) of undergraduate students on a range of professional courses in the Health Sciences Faculty in Hong Kong Polytechnic University, so similar courses to the context of this study; Kember et al. reported satisfactory reliability and validity indices; it was originally written in English; it has also been independently assessed in a number of studies, including Lethbridge et al. (2013) who conducted a comprehensive study comprised of six administrations of the questionnaire and reported that its psychometric properties were reproducible in their context (undergraduate nurses on four courses in Ontario, Canada), and importantly; the items included were intentionally devised to test four separate factors relating to reflection. A study by Kalk et al. (2014) using responses from trainee teachers confirmed the two factor structure of the relevant eight items, so it was valid to administer the two sub-scales; Reflection and Critical reflection, rather than the whole questionnaire. In conclusion, there was more evidence in favour of using the Reflection Questionnaire and for selecting two of the factors (eight items) for inclusion in the questionnaire studies presented in this thesis.

6.3 The Reflection Questionnaire

Kember et al. developed the Reflection Questionnaire (RQ) in order to assess the extent to which students on professional courses engaged in reflective thinking, given that a common aim of healthcare professional courses is to develop the ability to reflect on practice (Mann et al., 2009). Kember et al. reported that they developed their scale substantially based on Mezirow's (1991) descriptions of reflective and non-reflective actions, but also with regard to the work of Dewey and others. The RQ scale contains four constructs; Habitual action, Understanding, Reflection and Critical reflection.

Habitual action refers to learned routines that are used so frequently that they become automatic and require little conscious thought, such as riding a bicycle. Understanding refers to learning or thinking about something without appraising it by relating it to other situations, so a student might understand a concept without reflecting on its significance for personal or practical application. These two constructs aim to assess non- reflective thinking to differentiate these from two higher-order activities relating to reflection.

Reflection is attributed when learners assess their experience to evaluate their actions for future improvement as well as considering different possible solutions to problems presented in the learning activity; they may critique assumptions about the content or process of learning. Critical Reflection refers to learners being aware of why they think, feel or act as they do and their ability to question their own assumptions or beliefs which underpin their chosen action or widely accepted knowledge or ideas.

Kember et al.'s underpinning theoretical framework posits that these four sub-scales are related and are not mutually exclusive, but represent a hierarchical continuum with Critical reflection held to be the most profound level of reflective ability. Each of the sub-scales that represent these constructs has four items, so the whole instrument is comprised of 16 items. As Habitual action and Understanding were held to assess non-reflective thinking or actions, I decided to administer only the eight items for the two sub-scales which did relate to reflective actions as these were the most pertinent to the aims of this thesis.

As noted in Chapter 4, the questionnaire used in my two studies was comprised of 20 items (including the 12 items relating to Engagement with Feedback), and pragmatically, I wanted to avoid respondent fatigue and keep the self-completion questionnaire sufficiently short that it could be completed in approximately 10 minutes. Respondents were asked to indicate their level of agreement with all items on a five-point Likert scale ranging from 'definitely disagree' (score as 1),'disagree with reservations' (score as 2), 'Only to be used if a definite answer is not possible' (score as 3), 'agree with reservations' (score as 4), and 'definitely agree' (score as 5). There were no negatively scored items in the RQ sub-scales. Appendices 3 and 4 show the order that the eight items were presented in on the questionnaire and Table 6.1 below shows the items in their sub-scales.

Table 6.1: Reflection and Critical Reflection Sub-scale Items

Reflection
1. I sometimes question the way others do something and try to think of a better way.
3. I like to think over what I have been doing and consider alternative ways of doing it.
5. I often reflect on my actions to see whether I could have improved on what I did.
7. I often re-appraise my experience so I can learn from it and improve for my next performance.
Critical Reflection
2. During this course I discovered faults in what I had previously believed to be right.
4. As a result of this course I have changed the way I look at myself.
6. This course has challenged some of my firmly held ideas.
8. As a result of this course I have changed my normal way of doing things.

6.4 Two Administrations of the Reflection and Critical reflection items

Methods and Participants

The demographic characteristics of the respondents and their scores in the relevant end of year clinical examination, as well as evidence to show that these two convenience samples of students are representative of their wider populations are reported in Chapter 7 (Phase 1 administration) and Chapter 8 (Phase 2 administration). The methods of data collection and analysis are reported in Chapter 4 and discussed in more detail in Chapter 7 and 8. The numbers of respondents were 180 in Phase 1 and 233 in Phase 2.

Results of Two Administrations of the Reflection Questionnaire Sub-scales

The data from the two RQ sub-scales were subjected to exploratory factor analysis to provide potential evidence to establish the validity of the factors for the participants in these two studies.

Phase 1

The data from the Phase1 administration of the 8 items from the RQ scale were subject to principal components analysis (PCA) using IBM SPSS version 22. Prior to performing PCA, the suitability of the data for factor analysis was assessed to ascertain that factor analysis was a valid approach (Pallant 2013, p207). The correlation matrix showed many coefficients of .3 and above. The Kaiser-Meyer-Okin value was .733 and so above the recommended minimum value of .6 and Bartlett's Test of Sphericity reached statistical significance and thus supported the

factorability of the correlation matrix. PCA indicated the presence of two components with eigenvalues exceeding 1, which accounted for 36.5% and 17% of variance respectively. An inspection of the scree plot revealed a clear break after the second component. The two component solution explained a total of 53.5%of the variance. The two factors were weakly correlated (r=.295).

An oblique method of rotation, Direct Oblimin was performed to examine the dimensionality of the sub scales. This method was chosen because Kember et al. (2000) argued that the sub-scales are not mutually exclusive, but represent a hierarchical continuum and so a degree of correlation between factors was expected. The communality estimates for the items ranged between .301 and .685. The rotated solution indicated a number of strong loadings and all but one (Q3 'I like to think over what I have been doing and consider alternative ways of doing it') loaded substantially on only one component. However two of the items loaded onto the opposite factor than the one predicted. These were Q1 ('I sometimes question the way others do something and try to think of a better way') which loaded on Critical reflection and Q8 (As a result of this course I have changed my normal way of doing things) which loaded on Reflection.

Table 6.2:Phase 1 Pattern and Structure Matrix for PCA with Direct Oblimin Rotation of Two Factor Solution of 8 RQ Items

Pattern Coefficients			**Structure Coefficients**			**Communalities**	
	Component			Component			
Q	R	CR	Q	R	CR	Q	
7	**.860**	-.155	7	**.814**	.099	1	.511
5	**.818**	-.017	5	**.813**	.224	2	.626
8	**.662**	.113	8	**.695**	.308	3	.554
3	**.499**	**.424**	3	**.624**	**.571**	4	.446
2	-.197	**.826**	2	.047	**.768**	5	.662
1	-.067	**.732**	1	.149	**.712**	6	.301
4	.279	**.530**	4	.435	**.612**	7	.685
6	.147	**.487**	6	.290	**.530**	8	.495

Notes: Major loadings for each component are in bold.
Component names: R=Reflection, CR=Critical reflection

Phase 2

The data from the Phase 2 administration of the 8 items from the RQ scale were subject to principal components analysis (PCA). Prior to performing PCA, the suitability of the data for factor analysis was assessed to ascertain that factor analysis was a valid approach. The correlation matrix indicated that there were some coefficients of .3 and above. The Kaiser-Meyer-Okin value was .749 and so above the recommended minimum value of .6 and Bartlett's Test of Sphericity reached statistical significance and thus supported the factorability of the correlation matrix. PCA indicated the presence of two components with eigenvalues exceeding 1, which accounted for 34.2% and 17.7% of variance respectively. An inspection of the scree plot revealed a break after the second component. The two component solution explained a total of 51.9%of the variance. There was a medium strength correlation between the two factors (r=.312).

Direct Oblimin was performed to examine the dimensionality of the sub scales. The communality estimates for the items ranged between .241 and .645. The rotated solution indicated a number of strong loadings with all items only loading substantially on the one component that was theoretically predicted.

Table 6.3:Phase 2 Pattern and Structure Matrix for PCA with Direct Oblimin Rotation of Two Factor Solution of 8 RQ Items

Pattern Coefficients			**Structure Coefficients**			**Communalities**	
	Component			Component			
Q	R	CR	Q	R	CR	Q	
7	**.793**	-.062	5	**.802**	.284	1	.377
5	**.791**	.038	7	**.774**	.185	2	.241
3	**.709**	.005	3	**.711**	.226	3	.505
1	**.604**	.032	1	**.613**	.220	4	.551
8	-.111	**.821**	8	.145	**.786**	5	.645
6	.069	**.747**	6	.302	**.769**	6	.595
4	.061	**.721**	4	.286	**.740**	7	.602
2	.006	**.489**	2	.159	**.491**	8	.629

Notes: Major loadings for each component are in bold.
Component names: R=Reflection, CR=Critical reflection

The Cronbach's alpha reliability values from my administration of the two sub-scales are shown in Table 6.4 below. These are similar to those reported in Kember at al's (2000) study, and to other studies shown below.

Table 6.4 Comparison of Estimates of Reliability (internal consistency) for Reflection and Critical Reflection sub-scales from various studies involving students on professional courses

	Cronbach's Alpha	
Study	**Reflection**	**Critical Reflection**
Phase 1	.681	.623
Phase 2	.704	.662
Kember et al. (2000)	.631	.675
Lethbridge et al. (2013)	.684	.753
Kalk et al. (2014)	.77	.70
Tricio et al. (2015)	.702	.712

Note: For Phase 1 items Q1 and Q8 were allocated to their expected sub-scales

The mean scores with standard deviations for Phase 1 and 2 were comparable with those reported in various other studies.

Table 6.5: Comparison of mean score (out of 20) & Standard Deviations for Reflection and Critical Reflection sub-scales from various studies

Study	**R Mean (SD)**	**CR Mean(SD)**
Phase 1 (UG Medics Y3 & 4, N=180)	16.13 (2.51)	15.51 (2.74)
Phase 2 (UG Medics Y3, 4 & 5, N=233)	16.51 (2.45)	15.57 (2.76)
Kember (UG & PG Healthcare Professionals, N=303)	15.25 (2.21)	12.70 (2.82)
Lethbridge (UG Y3 Nurses, N=538)	16.59 (2.37)	15.32 (3.10)
Tricio (UG & PG Dentists N=324)	17.0 (1.80)	14.4 (2.0)

An independent-samples t-test was conducted to compare the Reflection scores for Phase 1 and Phase 2 respondents. There was no significant difference in the scores for Phase 1 (M=16.13, SD=2.51) and Phase 2 (M=16.51, SD= 2.45; t(411)=-1.56, p=.12 two-tailed, *d*=-0.15). The same test was conducted to compare the Critical reflection scores. Again, there was no significant difference in the scores for Phase 1 (M=15.51, SD=2.74) and Phase 2 (M=15.57, SD= 2.76; t(411)=-.203 p=.84 two-tailed, *d*=-0.02).

A paired samples t-test was conducted to ascertain if there was a difference between the mean scores for Reflection and Critical reflection for Phase 1 respondents. There was a statistically significant difference between the scores on these two sub-scales; Reflection (M=16.13, SD 2.51) and Critical reflection (M=15.51 SD=2.74), t(179)=3.09, p=.002 two tailed. The same test was conducted on the Phase 2 data which also indicated that there was a statistically significant difference between the scores on these two sub-scales; Reflection (M=16.51, SD2.45) and Critical reflection (M=15.57 SD=2.76), t(232)=4.76, p=<.001 two tailed.

6.5 Discussion

The PCA results for the Phase 1 data indicated some anomalies from those theoretically expected; two questions were factored into the opposite component to the one theoretically predicted; Reflection Q1 was classified as a Critical reflection item and Q8 as a Reflection item. Also Q3 loaded strongly on to both components. However, the results from the PCA of the Phase 2 data, which was a larger sample size, factorised the items as expected, so the Phase 1 anomalies may have been caused by the sample size (Costello and Osborne 2005, p3). Anomalies similar to the results in Phase 1 may have been present in other studies, potentially not published due to publication bias whereby positive results have a better chance of being published (Peplow, 2014).

The results for the estimates of internal consistency (Cronbach's alpha) for the two sub-scales were similar for the Phase 1 and Phase 2 administrations and were both comparable with those reported by Kember et al. (2000) and in the same range as those reported in three other studies.

The independent-samples t-test results indicated that the mean scores for the two administrations of both sub-scales were not significantly different which provides evidence that the instrument has temporal stability.

The paired samples t-test results suggest that there was a statistically different mean score for the Reflection score compared with the Critical reflection score for both administrations of the sub-scales. This finding fits with Kember et al.'s (2000)

conceptual description as it is expected that undergraduate students would use the Critical reflection domain less than the Reflective domain because 'Critical reflection requires a major change of perspective and alteration to deep-seated beliefs which is a difficult, lengthy and often painful process' (Kember et al. 2000, p391) and so undergraduates may not have the necessary exposure to clinical practice, and are not directly responsible for patient care which could present prompts for deeper, critical reflection. This point is also highlighted in Kember et al.'s (2000, p391) and Tricio et al.'s (2015, p118) studies, where the post-graduate students in their contexts scored more highly on the Critical reflection sub-scale than the undergraduate students and this was held to be because they had more clinical experience and responsibility, and were usually continuing in their professional practice whilst studying.

6.6 Conclusion

The RQ was developed to assess the extent to which students on professional courses engage in reflective thinking during their course, taking into account that not all thinking used whilst learning would be reflective in nature. The initial sub-scales were developed from a secure theoretical base and were piloted and revised after being subject to exploratory factor analysis and confirmatory factor analysis. The final version of the RQ was robustly tested by Kember et al. (2000) and subsequently by other authors, so the validity and reliability of the scale has been established in similar contexts to this one.

The internal consistency (reliability) of the results from the two RQ sub-scales over two administrations appears to be comparable to the results of other studies, and was consistent over the two administrations. Although exactly the same process for PCA was conducted on the RQ data in Phase 1 and Phase 2, the resulting factored sub-scales for Phase 1 were not as theoretically expected for two items.

6.7 Summary

The aim of this Chapter was to explore the reliability and validity of the two sub-scales of Kember et al.'s (2000) Reflection Questionnaire in this context by comparing the results from the Phase 1 and 2 administrations of the Reflection

and Critical reflection sub-scales with each other and with published data from other administrations of the RQ to ascertain if the results were as theoretically expected.

CHAPTER 7: STUDY 2 - DEVELOPMENT OF A SURVEY TO ASSESS ENGAGEMENT WITH FEEDBACK PHASE 1

7.1 Introduction

As described in Chapter 3, an outline conceptual framework based on self-regulated learning (SRL) theory (Nicol and Macfarlane-Dick, 2006, Sandars and Cleary, 2011) was developed which highlighted the roles of reflection on, and engagement with, written feedback from a summative, end of academic year clinical examination. A qualitative study was undertaken in order to substantiate this outline conceptual framework. Fifth year students, whose scores from their end of fourth year (Y4) clinical examination fell into one of three specified performance levels (top, borderline pass or fail), were invited to participate in the study. Eleven students volunteered to attend for a semi structured interview to explore their views and actions with regard to the feedback provided after the Y4 clinical examination. This study (reported in Chapter 5) indicated that the top performing respondents were intrinsically motivated and more effective than borderline and fail respondents at reflecting and acting on the feedback to improve their skills and knowledge for longer term clinical practice. The results of the interviews enabled the outline conceptual framework to be further developed. The final SRL conceptual framework (Chapter 5, Figure 5.3) highlighted the importance of reflection for students' self-assessment of their examination performance (before and after the feedback was released) as a key indicator of effective engagement with the feedback.

The next step in this thesis was to investigate how a wider range of students (i.e. larger numbers of respondents who were not sampled by specific performance categories) engaged with their feedback comments. The data for this study was collected via a self-completion survey instrument which included questions in order to explore if students' propensity to reflect on their learning in general or to engage with feedback was associated with the their level of performance in the clinical examination that generated the feedback.

As reported in Chapter 4, a review of the relevant literature suggested that there were no published survey instruments which related to student engagement with feedback that were appropriate for the aims of this study. Therefore, a set of statements was constructed based on the SRL conceptual framework and influenced by what the top performing students in the previous qualitative study had reported regarding how they had engaged with the personalised feedback. As reported in Chapter 6, in order to explore the students' propensity to reflect on learning (an important element of the SRL conceptual framework), two of four sub-scales from the Reflection Questionnaire (RQ) (Kember et al. 2000); Reflection and Critical reflection were chosen as these appeared to relate most closely to aims of this thesis.

The Phase 1 questionnaire comprised four questions on respondents' demographic background, two on their performance in their most recent end of year clinical examination, the four items from each of the two RQ sub-scales, and 12 items for Engagement with Feedback (EWF). Four of these 12 items were about students' thinking and actions after sitting the clinical examination and whilst waiting for the feedback to be released. If students had accessed their feedback, they were also asked to complete a further eight items on their thoughts and actions with regard to the feedback comments.

7.1.1 Aims of this study

As noted in Chapter 3, Sundre and Kitsantas (2004) claim that effective self-regulation is a characteristic that distinguishes between high and low performing students in higher education settings. Students who are effective at SRL should perform better in the clinical examination and so their scores from this might be viewed as a proxy measure for their SRL ability. I hypothesised that effective self-regulated learners (i.e. higher performers) will be better able to reflect on their performance in the clinical examination and are more likely to be motivated to access, critically reflect on and act on their feedback. The three main aims of this study were firstly to pilot a set of survey items to begin to ascertain if the initial three hierarchical Engagement with Feedback sub-scales; 'Reads', 'Thinks', 'Acts' were a valid and reliable instrument; secondly to confirm if a previously validated scale to assess propensity to reflect on learning activities had similar outcomes in

this context; and thirdly to ascertain if there was an association between examination performance, reflection on learning and/or engagement with feedback comments.

7.1.2 Research questions

The questions this study aims to address via data generated from the questionnaire are:

1. Do the proposed Engagement with Feedback sub-scales effectively assess student engagement with the feedback comments provided after their clinical examinations?
2. Is the level of student engagement with feedback associated with their performance level in their end of year clinical examinations?
3. Is the students' propensity to reflect on their learning in general associated with their performance level in their end of year clinical examinations?

7.2 Research design

7.2.1 Survey instruments

Engagement with Feedback (EWF) Sub-scales

The 12 item EWF scale had three sub-scales, 'Reads' (2 items), 'Thinks' (6 items) and 'Acts' (4 items). The scale was presented in a Likert format described in Chapter 4. Three of the questions were reversed scored (9, 17 and 20 in Table 7.1 below).

Reflection Questionnaire Sub-scales

The two RQ sub-scales used were 'Reflection' and 'Critical Reflection'. Each of these scales had four items presented in a Likert format described in Chapter 4. None of these items were reversed scored.

Questionnaire Design

The two sets of scales were administered as one questionnaire, divided into three sections. Section A included the RQ scale items. Section B related to thoughts and actions before the feedback was available. Section C related to thoughts and actions after accessing the feedback. The version of the questionnaire completed by the students did not include sub-scale headings. Table 7.1 shows the items

under their sub-scale headings and with the number order they were presented in. The full survey is shown as Appendix 3.

Table 7.1: Phase 1 Questionnaire Sub-scale Items

Reflection	
1.	I sometimes question the way others do something and try to think of a better way.
3.	I like to think over what I have been doing and consider alternative ways of doing it.
5.	I often reflect on my actions to see whether I could have improved on what I did.
7.	I often re-appraise my experience so I can learn from it and improve for my next performance.
Critical Reflection	
2.	During this course I discovered faults in what I had previously believed to be right.
4.	As a result of this course I have changed the way I look at myself.
6.	This course has challenged some of my firmly held ideas.
8.	As a result of this course I have changed my normal way of doing things.
Reads	
13.	I looked at all of my individual feedback comments.
18.	I read my individual feedback comments as soon as I was able to access them.
Thinks	
9.R	I did not think about how I had done on each station because it was too late to change anything. [Before FB was available]
11.	I asked other students how they had answered questions to help me to work out how I had done on each station. [Before FB was available]
12.	I thought to myself about how I had done on each station. [Before FB was available]
14.	I thought about how I could use my individual feedback comments to improve my knowledge or skills.
16.	I thought about whether I agreed or disagreed with each of my individual feedback comments.
20. R	Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them.
Acts	
10.	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills. [Before FB was available]
15.	I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.
17. R	As I wanted to start this year fresh I am not planning to act on the suggested areas to improve in my individual feedback comments.
19.	I took specific action to improve my knowledge or skills as suggested in my individual feedback comments

Notes: R = Question results reversed prior to analysis

7.2.2. Participant recruitment

Participants were a convenience sample of medical students at University of Birmingham, UK who were in 2015-16 cohort of Year 4 or 5 and who were attending three specific lectures held in the School of Medicine

7.3 Results

7.3.1 Participants

68 Year 4 students (19% of the total population) and 112 Year 5 students (31% of total population) volunteered to complete the questionnaire with respect to the feedback they received after their end of year clinical examinations in April 2015 (the previous academic year). It is not known what proportion of the total cohort was present when the questionnaire was handed out. The questionnaire was circulated to Y4 during a plenary lecture on 4th February 2016, when all students were expected to be present. It was circulated to Y5 during two lectures (27th January and 17th February 2016), at each of which subsets of approximately 90 randomly allocated students should have been present. This subset of 180 students represents 50% of the total Y5 population. Therefore with respect to the 112 Y5 responses, the response rate would be 62% of the students requested to complete the survey, if all students were present as expected.

A total of 180 students (68 in Y4 and 112 in Y5) completed the questionnaire out of a possible total cohort of about 720. Although the respondents were drawn from a convenience rather than a random sample, the chi square goodness of fit results show that, except for the responders by type of course, the demographic characteristics of the convenience sample were not significantly different to a random sample and so for these characteristics, the sample is broadly representative of the total population of students. The result for responders by course shows that significantly more GEC/Dental students responded than would be expected in a random sample. These students are older than the average for the main course and this is shown in the higher than expected number of responders in the 26-30 and 31+ age bands, although these are not significantly higher than expected in a random sample. Demographic data for respondents is shown in Tables 7.2 and 7.3 below.

Table 7.2: Y4 & Y5 Respondent Demographic Data Compared to Total Cohort

	Responders N	Total Population N	Resps as % of total pop	χ^2 For Goodness Of Fit
Course				
Main	147	637	23.1	There is a significant difference in the proportion of Main and GEC respondents as compared to the total cohort population, χ^2 (1, n=177) =10, p=0.001
GEC/Dentists	30	83	36.1	
Missing	3			
Domicile				
Home/EEC	168	669	25.1	No significant difference in the proportion of Home/EEC and International respondents as compared to the total cohort population, χ^2 (1, n=177) =1.742, p=0.187
International	9	51	17.6	
Missing	3			
Gender				
Female	119	445	26.7	No significant difference in the proportion of Male and Female respondents as compared to the total cohort population, χ^2 (1, n=180) =2.48, p=0.115
Male	61	275	22.2	
Age bands				
21 or less	19	84	22.6	No significant difference in the proportion of respondents in each of 4 age bands as compared to the total cohort population, χ^2 (3, n=180) =3.705, p=0.295
22-25	141	573	24.6	
26-30	18	57	31.6	
31+	2	6	33.3	

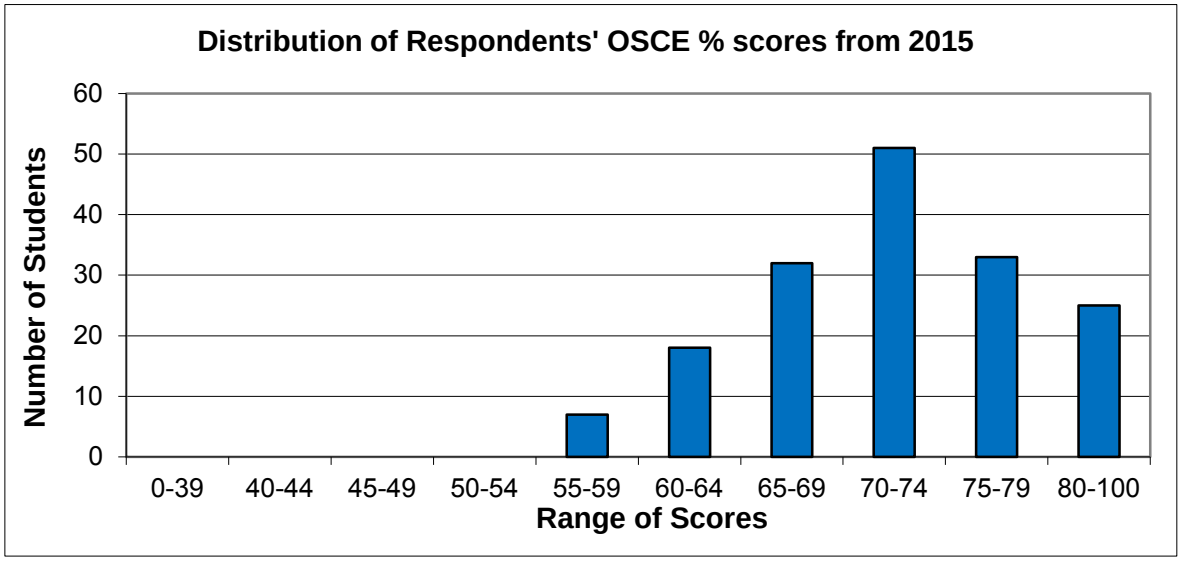
NOTE: GEC = Graduate Entry Course, graduate entry students undertake a separate first year of study and then join the main cohort in year 3. Dentists have a primary qualification in dentistry and directly enter the main course in year 3.

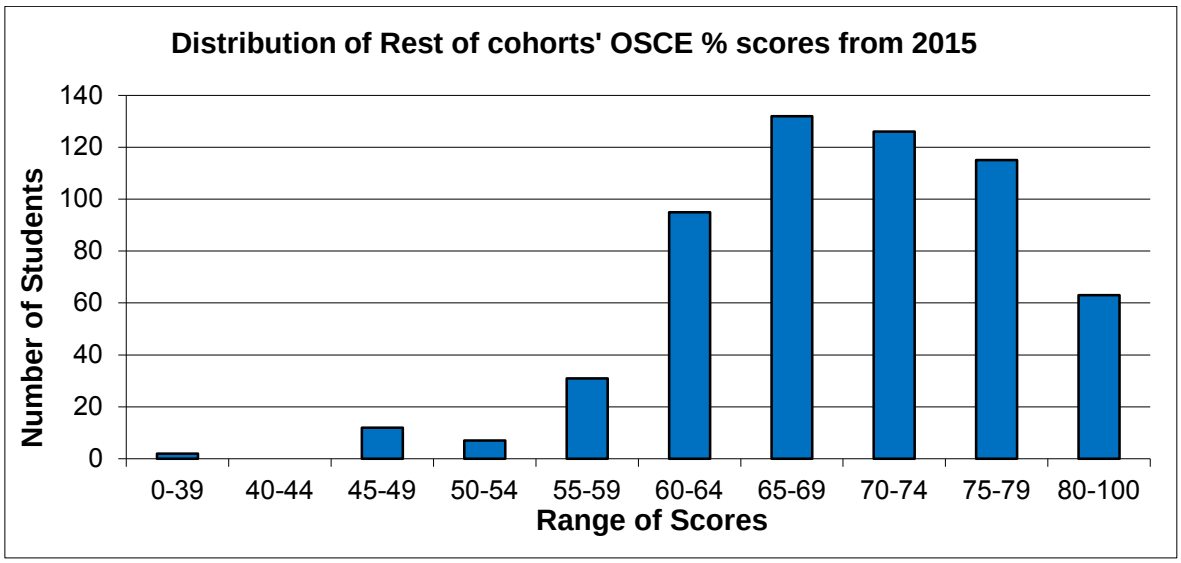
Table 7.3: Breakdown of Respondents' Demographic Characteristics by Year

	Responders N	Total Population N	Resps as % of total pop	Responders N	Total Population N	Resps as % of total pop
	Y4			**Y5**		
Course						
Main	52	320	16.3	95	317	30.3
GEC/Dentists	14	39	35.9	16	44	36.4
Missing	2			1		
Domicile						
Home/EEC	66	341	19.4	102	328	31.1
International	1	18	5.6	8	33	24.2
Missing	1			2		
Gender						
Female	48	218	22.0	71	227	31.3
Male	20	141	14.2	41	134	30.6
Age bands						
21 or less	18	83	21.7	1	1	100.0
22-25	41	252	16.3	100	321	31.2
26-30	8	23	34.8	10	34	29.4
31+	1	1	100.0	1	5	20.0

The range of scores achieved by the respondents in the end of year clinical examination (OSCE) roughly approximates the range of scores achieved by the total population. The convenience sample does not include any students who failed the first sit of the OSCE. Only two Y4 students and eight Y5 students in the total population were in this position. Similarly, there were no respondents in the 50-54% score range, but there were only one Y4 and two Y5 students in the total population. All the other performance categories were well represented. The distribution of the OSCE scores for the sample and the rest of the population are shown in Figure 7.1 below.

Figure 7.1 Distribution of OSCE Percentage Scores for Sample and Rest of Population Cohorts





7.3.2 Data collation

The questionnaire was designed to be read by an optical mark recognition (OMR) machine. The completed questionnaires were scanned using Remark Office OMR (version 8) software, and a document scanner (Fujitsu fi-5530-C2). The resulting data were initially imported into Microsoft Excel 2010 spreadsheets for error checking, sorting and coding of variables into data suitable for importing into IBM SPSS Statistics 22 for full analysis. For the purposes of this study, the results from Y4 and Y5 students were combined and treated as one set of 180 respondents. Missing data was dealt with via the 'exclude cases pairwise' option

in SPSS so that cases are excluded if they are missing the data required for a specific statistical analysis. Cases are still included in any of the statistical analyses for which they have provided the necessary information (Pallant 2013, p131). Appendix 6 details the missing data per respondent.

7.3.3 Individual Item Statistics

The full range of scores was not selected for four of the 20 questions. The minimum mean score was 3.41 and the maximum was 4.71 out of five. (Table 7.4)

Table 7. 4 Individual item statistics

Question No	Range	Mean score	Std. Deviation	N Respondents
1	1-5	4.06	0.83	180
2	1-5	4.20	0.73	177
3	2-5	4.03	0.83	179
4	1-5	4.10	0.99	179
5	1-5	4.12	0.85	180
6	1-5	3.55	1.06	180
7	2-5	3.96	0.89	179
8	1-5	3.77	0.99	179
9^R	1-5	3.63	1.33	179
10	1-5	3.41	1.24	179
11	1-5	3.57	1.39	179
12	1-5	4.37	0.95	179
13	1-5	4.71	0.71	167
14	2-5	4.18	0.91	167
15	1-5	3.81	1.06	167
16	2-5	4.31	0.81	167
17^R	1-5	4.04	0.96	166
18	1-5	4.22	1.11	167
19	1-5	3.42	1.11	166
20^R	1-5	3.46	1.19	166

Note: ^{R}= Question scores reversed prior to analysis

7.3.4 Initial sub-scale statistics

As expected, there were a number of significant correlations between the items which comprise the two RQ sub-scales (Reflection items1, 3, 5, 7 and Critical Reflection items 2, 4, 6, 8) although not every sub-scale item correlates significantly with the other three in the sub-scale (e.g. item 7 could be expected to correlate with item 1, and item 2 with item 8).

With regards to the EWF items, again there were a number of significant correlations between items on the same hypothesised sub-scales (Reads items 13 and 18; Thinks items 9, 11, 12, 14, 16 and 20; Acts items 10, 15, 17, 19) and between items on other sub-scales (Table 7.5).

As expected, there were a number of significant correlations between sub-scales on the same scale (Table 7.6). The strongest correlations between the sub-scales were; Thinks with Reads (0.539, $p<0.01$), Thinks with Acts (0.466, $p<0.01$) and Reflection with Critical Reflection (0.462, $p<0.01$). There were also significant correlations between the EWF sub-scales and the Reflection sub-scale. The strongest of these was Reflection with Thinks (.232, $p<0.01$). The strongest correlation with the OSCE scores were the Reads sub-scale (.308, $p<0.01$) and the 12 EWF items combined as one scale (.205, $p<0.01$).

Table 7.5: Pearson's Product-Moment Correlation Coefficient for the 20 RQ and EWF Items

Item	1	2	3	4	5	6	7	8	9[R]	10	11	12	13	14	15	16	17[R]	18	19	20[R]
1	-																			
2	**.403****	-																		
3	**.384****	**.330****	-																	
4	**.191***	**.283****	**.333****	-																
5	**.155***	.088	**.500****	**.273****	-															
6	**.170***	**.204****	**.182***	**.400****	**.160***	-														
7	.095	.060	**.398****	**.149***	**.558****	.128	-													
8	**.160***	.067	**.305****	**.419****	**.354****	**.285****	**.423****	-												
9[R]	-.070	.010	.052	.003	.105	-.118	**.246****	.038	-											
10	.003	-.099	-.052	-.038	.064	-.081	.094	.085	**.184***	-										
11	-.026	-.055	.066	-.029	.102	-.068	.068	**.148***	**.382****	**.171***	-									
12	.049	.044	.147	.073	.088	-.004	**.148***	**.151***	**.445****	**.151***	**.484****	-								
13	.081	.029	.127	.014	.125	-.121	.114	.059	**.235****	.067	**.191***	**.249****	-							
14	.104	-.082	.060	-.079	.138	-.134	**.157***	**.238****	.113	**.317****	.127	**.198***	**.353****	-						
15	.134	.071	**.167***	.006	.138	-.043	.067	**.198***	.053	**.234****	.112	.129	**.181***	**.544****	-					
16	.106	.097	**.184***	.114	**.156***	-.003	**.236****	**.199***	**.191***	.066	.017	.128	**.458****	**.275****	**.211****	-				
17[R]	-.012	.047	.028	-.061	.017	-.075	-.005	-.054	.002	-.011	-.005	.063	**.189***	**.290****	**.181***	.093	-			
18	-.014	.031	**.153***	.013	.141	.033	**.183***	.076	**.315****	.139	**.179***	**.429****	**.432****	**.312****	**.185***	**.308****	.112	-		
19	.012	-.078	.089	-.042	**.277****	.042	**.328****	**.217****	.106	**.404****	**.201****	**.180***	**.198***	**.517****	**.430****	**.307****	.101	**.361****	-	
20[R]	.038	.016	.141	-.023	**.173***	-.016	**.211****	.130	**.165***	**.201****	.067	.142	**.292****	**.397****	**.379****	**.367****	**.374****	**.208****	**.449****	-

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

[R] Question scores reversed prior to analysis

Table 7.6: Pearson's Product-Moment Correlation Coefficient for RQ & Initial EWF Scales and Sub-scales with OSCE Scores

Scales	Reflection	Critical Reflection	Reads	Thinks	Acts	RQ	EWF
Reflection	_						
Critical Reflection	**.462****	_					
Reads	**.186***	.040	_				
Thinks	**.232****	.059	**.539****	_			
Acts	**.155***	-.016	**.323****	**.466****	_		
Reflection Questionnaire	(.841**)	(.867**)	.133	**.170***	.079	_	
Engagement With Feedback	.**228****	.015	(.669**)	(.701**)	(.815**)	.139	_
OSCE Score	**.170***	**.173***	**.308****	**.165***	.090	**.199***	**.205****

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)
Note: Brackets highlight self-correlation scores of scales with their sub-scales

Cronbach's alpha score was calculated to assess the reliability (internal consistency) of the scales and sub-scales. The values for the three scales from the 12 EWF items were lower than the normally considered acceptable cut off score of 0.7 (Pallant 2013, p104). Cronbach's alpha is usually positively influenced by the number of items included in the calculation. However, the value of the four questions in the 'Thinks' sub-scale was 0.633, compared to the six questions in the 'Acts' sub-scale of 0.538. The results for the eight items from the two Reflection Questionnaire sub-scales considered as one scale called RQ and the 12 items from the three Engagement With Feedback sub-scales considered as one scale called EWF are shown to explore if this might be a fruitful approach. However, despite the number of items per scale, the Cronbach's score per scale does not increase substantially, although they are both increased to above the normally considered acceptable cut off score of 0.7.

Table 7.7: Estimates of Reliability (internal consistency) and descriptive statistics for Scales and Sub Scales

Scale/Sub-scale	Cronbach's Alpha	Mean	SD	Range
Reflection	.681	4.04	0.61	3.50
Critical Reflection	.623	3.90	0.65	3.25
Reads	.563	4.46	0.77	3.50
Thinks	.633	3.91	0.70	3.67
Acts	.538	3.63	0.77	4.00
8 RQ items (Q1-8)	.748	3.97	0.54	3.00
12 Engagement with Feedback items (Q9-20)	.771	3.79	0.54	2.50

7.3.5 Factor Analysis

Assessment of the suitability of the data from the EWF items for factor analysis

As recommended by Pallant (2013, p 134), preliminary analyses were conducted to ensure there was no violation of the assumptions of normality, linearity and scores per each question are evenly spread (homoscedasticity). Analysis of results for individual EWF items indicated that although the full range of scores was used for 10 items out of 12, the lowest mean score for the EWF items is 3.41 and the highest 4.71 out of a maximum possible score of 5 (Table 7.4 above). Therefore scores per question are not normally distributed. Pallant (2013, p59) quotes Tabachnick and Fidell's (2013, p80) judgement that 'with reasonably large sample sizes, skewness will not make a substantive difference'.

Pallant (2013, p190) notes there are differences of opinion between authors about minimum sample size, not just regarding the number of cases, but also the ratio of participants to items. Her general recommendation is 'the larger the better'. Robson (2011, p442) suggests there should be a minimum of five times the number of participants to the number of variables. The sample maximum in this study was 180 and because some questions were not answered by up to 14 respondents, the sample minimum was 166. This is a reasonable number, and as there are 12 questions, there are a minimum of 13 respondents per question. Finally, the strength of relationship between the items should be considered. Pallant (2013, p190) suggests that if there are few correlations above 0.3 then factor analysis may not be appropriate.

Principal Components Analysis (PCA)

The 12 items in the EWF scale were subject to principal components analysis (PCA) using IBM SPSS version 22. Prior to performing PCA, the suitability of the data for factor analysis was assessed to ascertain that factor analysis was a valid approach (Pallant 2013, p207). The correlation matrix showed many coefficients of .3 and above. The Kaiser-Meyer-Okin value was .756 and so above the recommended minimum value of .6 and Bartlett's Test of Sphericity reached statistical significance and thus supported the factorability of the correlation matrix. The three components with eigenvalues exceeding 1 were extracted. These accounted for 30.6%, 14.3% and 10.7% of variance respectively. An inspection of the scree plot revealed a gentle break after the third component. The three component solution explained a total of 55.6%of the variance.

An oblique method of rotation, Direct Oblimin was performed to examine the dimensionality of the sub scales. This method was chosen because it was hypothesised that the EWF sub- scales were not mutually exclusive, i.e. it is logical that a student could give positive answers to lower level engagement questions and higher level engagement ones. The communality estimates for 11 of the items ranged between .642 and .507 but Q17 was an outlier at .318. The rotated solution showed a number of strong loadings and most variables loading substantially on only one component. However some of the items in component number three also loaded strongly on the other two scales and did not make sense conceptually as a sub-scale.

The version of the sub-scales that was most interpretable with reference to the conceptual framework was a PCA with 10 items (excluding reversed score items 17 and 20). The Kaiser-Meyer-Okin value was .740, so above the recommended minimum value of .6 and Bartlett's Test of Sphericity reached statistical significance. PCA suggested the presence of three components with eigenvalues exceeding 1, which accounted for 33%, 15.8% and 12.4% of variance respectively. An inspection of the scree plot revealed a clear break after the third component. The three component solution explained 61.2% of the variance cumulatively. The variance communality estimates ranged from .671 to .535. The rotated solution indicated a number of strong loadings and most variables loading

substantially on only one component. All three components could be accounted for within the conceptual framework (Table 7.8).

Table 7.8: Pattern and Structure Matrix for PCA with Direct Oblimin Rotation of the Three Factor Solution of 10 Engagement With Feedback Items

	Pattern Coefficients Component				Structure Coefficients Component			Communalities	
Q	H	T	L	Q	H	T	L	Q	
19	**.766**	.024	-.107	19	**.795**	.213	-.285	9	.576
15	**.752**	-.126	-.083	14	**.784**	.149	-.402	10	.535
14	**.742**	-.058	-.244	15	**.742**	.056	-.233	11	.624
10	**.675**	.201	.302	10	**.651**	.300	.114	12	.667
12	.004	**.789**	-.111	12	.206	**.809**	-.247	13	.662
11	.081	**.783**	.156	11	.221	**.774**	.003	14	.671
9	-.094	**.739**	-.146	9	.105	**.743**	-.252	15	.570
16	.104	-.092	**-.780**	13	.222	.282	**-.800**	16	.635
13	.014	.146	**-.772**	16	.260	.064	**-.787**	18	.543
18	.119	.350	**-.542**	18	.320	.470	**-.629**	19	.644

Notes: Major loadings for each component are in bold.
Component names: T= Thinking for P/F, L= Lower Level Engagement, H= Higher Level Engagement.

The estimates of internal reliability (Cronbach's Alpha) for these three components were calculated and were around the accepted cut-off point of .70 (Table 7.9).

Table 7.9: Estimates of Reliability (internal consistency) and descriptive statistics for 3 Engagement with feedback factors

Scale	Cronbach's Alpha	Mean	SD	Range
Thinking for Pass/Fail (Q9, 11, 12)	.70	3.86	0.97	4.00
Lower Level Engagement (Q13, 16, 18)	.67	4.41	0.68	3.33
Higher Level Engagement (Q10, 14, 15, 19)	.73	3.66	0.85	4.00
EWF 10 items	.76	3.93	0.63	3.40

The resulting three sub-scales broadly kept together the items from the initial sub-scales, but with some interesting differences. The first sub-scale named Thinking for Pass/Fail included three items which related to thinking about performance in terms of whether the student had passed or failed the clinical examination overall, before the feedback was released. The second sub-scale, named Lower Level Engagement included three items relating to reading and thinking about each of

the feedback comments, so described accessing and cognitively engaging with the comments. The third sub-scale, named Higher Level Engagement included the items from the previous Acts sub-scale along with Q14 which had originally been classified as a Thinks item, but described thinking about action (Table 7.10).

Table 7.10: Phase 1 Questionnaire Factored Sub-scale Items

Thinking for Pass/Fail	
9.R (T)	I did not think about how I had done on each station because it was too late to change anything. [Before FB was available]
11. (T)	I asked other students how they had answered questions to help me to work out how I had done on each station. [Before FB was available]
12. (T)	I thought to myself about how I had done on each station. [Before FB was available]
Lower Level Engagement	
13. (R)	I looked at all of my individual feedback comments.
16. (T)	I thought about whether I agreed or disagreed with each of my individual feedback comments.
18. (R)	I read my individual feedback comments as soon as I was able to access them.
Higher Level Engagement	
10. (A)	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills. [Before FB was available]
14. (T)	I thought about how I could use my individual feedback comments to improve my knowledge or skills.
15. (A)	I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.
19. (A)	I took specific action to improve my knowledge or skills as suggested in my individual feedback comments
Deleted	
20. R (T)	Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them.
17. R (A)	As I wanted to start this year fresh I am not planning to act on the suggested areas to improve in my individual feedback comments.

Notes: R = Question results reversed prior to analysis.
Letters in brackets refer to the original sub-scale the item was in (R=Reads, T=Thinks, A=Acts)

Correlation analysis indicated positive, significant correlations between most of the individual EWF items. Where there was no significant correlation, the relevant items belonged in different sub-scales (Table 7.11).

Table 7.11: Pearson's Product-Moment Correlation Coefficient for 10 Engagement With Feedback items

Items	Q9^{R}T	Q11 T	Q12 T	Q13 L	Q16 L	Q18 L	Q10 H	Q14 H	Q15 H	Q19 H
Q9^{R}T	_									
Q11 T	**.382****	_								
Q12 T	**.445****	**.484****	_							
Q13 L	**.235****	**.191***	**.249****	_						
Q16 L	**.191***	.017	.128	**.458****	_					
Q18 L	**.315****	**.179***	**.429****	**.432****	**.308****	_				
Q10 H	**.184***	**.171***	**.151***	.067	.066	.139	_			
Q14 H	.113	.127	**.198***	**.353****	**.275****	**.312****	**.317****	_		
Q15 H	.053	.112	0.129	**.181***	**.211****	**.185***	**.234****	**.544****	_	
Q19 H	.106	**.201****	**.180***	**.198***	**.307****	**.361****	**.404****	**.517****	**.430****	_

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Notes: ^{R}= scores reversed prior to analysis. Scale name codes are included next to question numbers: T=Thinking for P/F, L= Lower level engagement, H= Higher level engagement.

As shown in Table 7.12, correlation analysis of the scales and factored EWF sub-scales indicated positive, small to medium but significant correlations between the three sub-scales from the 10 EWF items, which is to be expected as these are hypothesised to be measuring related concepts. There were similarly sized positive correlations between the Reflection sub-scale and the LLE and HLE sub-scales. A number of the scales and sub-scales have a small but positive significant correlation with the OSCE scores. The largest of these is the LLE sub-scale (0.293, $p<0.01$).

Table 7.12: Pearson's Product-Moment Correlation Coefficient for Scales and Factored Sub-scales with OSCE Scores

Sub-scale	Reflection	Critical Reflection	Thinking P/F	Lower Level Engage	Higher Level Engage	RQ	EWF
Reflection	_						
CR	**.462****	_					
Thinking P/F	.139	.027	_				
LLE	**.236****	.089	**.353****	_			
HLE	**.178***	-.003	**.237****	**.376****	_		
RQ	(.841**)	(.867**)	.100	**.189***	.100	_	
EWF	**.238****	.050	(.742**)	(.705**)	(.743*)*	**.168***	_
OSCE Score	**.170***	**.173***	.150	**.293****	.067	**.199***	**.175***

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Note: Brackets highlight self-correlation scores of scales with their sub-scales

Table 7.13: Percentage of Phase1 Responders Who Agreed or Strongly Agreed With the Item Statements with items shown in relevant sub-scales

	N	Item	% of respondents who agreed or strongly agreed
Kember Reflection	1	I sometimes question the way others do something and try to think of a better way.	88
	3	I like to think over what I have been doing and consider alternative ways of doing it.	86
	5	I often reflect on my actions to see whether I could have improved on what I did.	87
	7	I often re-appraise my experience so I can learn from it and improve for my next performance.	80
Kember Critical Reflection	2	During this course I discovered faults in what I had previously believed to be right.	91
	4	As a result of this course I have changed the way I look at myself.	82
	6	This course has challenged some of my firmly held ideas.	63
	8	As a result of this course I have changed my normal way of doing things.	75
Thinking for Pass/ Fail Lower level EWF	9	I did not think about how I had done on each station because it was too late to change anything. [Scores NOT reversed][Before FB was available]	30
	11	I asked other students how they had answered questions to help me to work out how I had done on each station. [Before FB was available]	70
	12	I thought to myself about how I had done on each station. [Before FB was available]	91
	13	I looked at all of my individual feedback comments.	96
	16	I thought about whether I agreed or disagreed with each of my individual feedback comments.	92
	18	I read my individual feedback comments as soon as I was able to access them.	82
Higher level EWF	10	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills. [Before FB was available]	63
	14	I thought about how I could use my individual feedback comments to improve my knowledge or skills.	85
	15	I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.	72
	19	I took specific action to improve my knowledge or skills as suggested in my individual feedback comments.	58
Deleted after FA	17	As I wanted to start this year fresh I am not planning to act on the suggested areas to improve in my individual feedback comments.	9
	20	Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them.	29

7.3.6 Answers to Research Questions

1. Do the proposed Engagement with Feedback sub-scales effectively assess student engagement with the feedback comments provided after their clinical examinations?

With regard to the first research question, the data from this study shows that the originally proposed sub-scales were not as effective at assessing a hierarchical approach to assessing engagement with feedback compared to the sub-scales produced after factor analysis.

2. Is the level of student engagement with feedback associated with their performance level in their end of year clinical examinations?

With regard to the second research question, the EWF sub-scale Lower Level Engagement is weakly, but significantly associated with performance level in the end of year clinical examinations.

3. Is the students' propensity to reflect on their learning in general associated with their performance level in their end of year clinical examinations?

With regard to the third research question, both the Reflection and Critical reflection sub-scales are weakly but significantly associated with performance level in the end of year clinical examinations.

7.4 Discussion

To arrive at the three sub-scale solution described above, analyses were performed with 10, 11 and 12 of the EWF items in two or three factor solutions. Including the reverse scored questions 17 and 20 made the resulting scales difficult to explain with reference to the conceptual framework, so although the Cronbach's alpha coefficient slightly improved with these questions included, it made more sense conceptually to exclude them**.** This phenomenon is reported in Swain et al. 'Despite the potential benefits of reversing scale items, some researchers have expressed concerns about the practice. These concerns stem from evidence linking item reversal to problems such as unexpected factor structures, diminished scale reliabilities and confounds in cross-cultural research' (2008 p116).

Overall, the sub-scales produced by the factor analysis of the 10 EWF items were an improvement on the initially proposed sub-scales. The original Reads sub-scale only had two variables which is an insufficient number. Pallant advises 'Ideally we would like three or more items loading on each component' (2013, p200). After factor analysis, each sub-scale had at least three items. The Cronbach's alpha values for the three new sub-scales from the 10 EWF items were improved from the original sub-scales comprising 12 items, with values for each increasing to around .7. The factor analysis enabled the conceptualisation of a more refined hierarchy of engagement compared to the initially proposed hierarchy of Reads, Thinks, and Acts. The new hierarchy of engagement ranged from Thinking for Pass/Fail (TPF) which encompassed thinking about performance in the examination before feedback was released in order to assess the likelihood of having passed or failed the examination, rather than in order to act to improve perceived deficits; to Lower Level Engagement (LLE) which described cognitively engaging with the feedback comments and Higher Level Engagement (HLE) which included thinking about acting or taking action to improve deficits in knowledge of skills.

A number of sub-scales had a small but statistically significant association with the OSCE scores, the largest of which was LLE. In terms of the hypothetical constructs, HLE, had been predicted to be more positively associated with OSCE scores, but did not correlate significantly. The items in the EWF scales may not have been effective at measuring the underlying construct of levels of engagement with feedback, or my hypothesis that higher scoring students would engage more effectively with the feedback was incorrect. Interestingly, the LLE and HLE sub-scales (but not the theoretically lower construct, TPF) had a similar small but positive association with the Reflection sub-scale, intimating that the conceptual framework correctly highlights the link between reflection, thinking and acting on the feedback. However, there was no association between HLE and Critical reflection as predicted in the hierarchical approach in these two scales.

Although many of the correlations reported in the relevant tables above are highlighted as being significant, Pallant cautions (2013, p129) 'With large samples,

even quite small correlation coefficients can reach statistical significance'. Pallant (2013, p139) suggests that Pearson's r can be classified as small r=.10 to.29, medium r= .30 to .49 and large r = .50 to 1.00, although the actual relevance of the size of the correlation depends on the research area and the results reported in related studies. It should be highlighted that the significant positive correlations between sub-scales and OSCE scores shown in Table 7.12 are classified as small.

The reliability and validity of the RQ sub-scales are discussed in Chapter 6 and these indicate that the results from these sub-scales are reliable and valid in this context. There is a weaker than theoretically expected association between the Reflection and Critical Reflection sub-scales with the level of performance in the OSCE. Grant et al. state that 'Studies have shown that greater reflective activities in learning is correlated with better performance in examinations' (2006, p380). However, they undertook a study which involved teaching volunteer participants in the third year of a UK medicine course how to reflect on learning activities and found that this had no significant effect on the end of year examination scores for the intervention group of 20 students compared to the rest of the cohort of 212 students.

Respondents

As with every other self-reporting survey, the respondents, may have been affected by what Streiner et al. (2015, p106) term 'social desirability' and so may have consciously or unconsciously wanted to appear to have done the 'right thing' and over-reported their engagement or reflection levels. It is also the case that poorer performing students do not necessarily realise that they are doing something poorly (Boud et al. 2015, p50) and so self-report that they do things more effectively than in actuality and this may inflate some of the reflection or engagement with feedback scores for this group.

Missing data: Respondents who did not read their feedback

13 students (seven Y4 and six Y5; 7.2% of total respondents) marked the lozenge to indicate that they had not read their feedback. Of these, four of the Y4 students

noted in the free text comments box that they had intercalated in the academic year following their Y3 OSCE 2014. As personalised feedback was only introduced for Y3 in 2014-15, they were not able to complete items 12 to 20. This fact had been overlooked when designing the questionnaire. It may be the case that the other nine students also intercalated or temporarily withdrew for other reasons, and so would not have received this feedback (if now in Y4) or may not have seen the relevance of accessing it as they would not be studying medicine for a year (e.g. if now in Y5). The Phase 2 version of the questionnaire will include a question to ask for data on this. The answers to the other questions from this group were retained for use in the analyses where they could be included via the 'pairwise' option in IBM SPSS. These students were mainly in the higher group of OSCE scores as would be expected for intercalating students. This may have had an effect on the results for the hypothesis that lower performing students would not engage with the feedback, but this was a relatively small number of students and the majority indicated that they had read the feedback, whatever their OSCE score.

Free-text feedback comments on the questionnaire

Respondents were invited to make free text comments about whether they had any difficulty with understanding or completing any of the demographic or Likert scale questions. It was decided to use the Reflection questionnaire Likert scale descriptors for the Engagement with feedback questions although they were not felt to be intuitive because on balance, it was held to be better to administer the Reflection Questionnaire items as faithfully to the original as possible and keeping to the same scale descriptors for all sections of the questionnaire seemed more logical. Five out of 112 Y5 students commented in the free text section that they found these descriptors confusing. There were very few other comments about the Reflection items or the Engagement with feedback items

Changes to the Phase 1 Questionnaire

The feedback comments from Y4 respondents regarding the wording of individual demographic data questions were used to update the version subsequently administered to Y5 in Phase 1. For example 'home' was changed to 'UK' and 'overseas' to 'international' student to reflect more commonly understood

nomenclature. To make instructions more clear, students were asked to round their OSCE score to the nearest whole number when indicating the relevant OSCE score range.

Planned changes for the Phase 2 Questionnaire

In Phase 1, the respondents were asked to report their OSCE scores in bands, which were converted to a score of 1-8 for the analyses, so this lack of sensitivity may have had an effect on the resulting correlations between the OSCE scores and the respondents' scores for reflection or engagement with feedback. In the next study, respondents will be asked to record their actual score to improve the granularity of this variable. The question on the number of stations passed will be deleted as this was initially thought to be another source of data to support the hypothesis that there would be an association between more able students' responses to the RQ and EWF scale scores. In practice it was not used as the number of stations changes per year of study and so is not useful in comparisons.

In Phase 2, respondents will be given the option of providing their student registration number or completing the demographic data anonymously. This will save respondents time and their ID number can be linked to official records so the data will be verified and potentially more accurate. The Phase 2 questionnaire will also include a question on whether the respondent intercalated or took leave of absence in 2015-16 as this could be a reason for not accessing the feedback.

For the reverse scored questions, (9, 17, and 20), a rationale was provided for why students might not do or think that (e.g. Q20 'Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them') and Clark and Watson (1995, p312) caution against this complex or double-barrelled approach as responders might have had another reason for not thinking about the comments. Although Q9 was effectively factored into the TPF sub-scale, the inclusion of Q17 and Q20 in the PCA did not lead to sub-scales that were interpretable with reference to the conceptual framework. In the next study, all three negatively phrased questions will be replaced with similar but positively phrased questions.

It was planned to administer the Phase 2 questionnaire to the 2016-17 cohorts of Y4 and Y5 students shortly after they commenced the academic year, and the outgoing 2015-16 Y5 students, just before they graduated, in order to increase the chances of recruiting a representative and reliable sample of students to administer the questionnaire with the updated version of the EWF scale and the two RQ sub-scales.

7.5 Conclusion

The respondent population was a reasonably large sample and was representative of the total population. As shown in Table 7.13, the majority of students reported that they agreed or strongly agreed with the items relating to reading and thinking about their feedback comments. The percentage of students who agreed or strongly agreed with the items relating to action was lower. This resonates with the findings in the previous study presented in Chapter 5, where the majority of students reported they had read the comments and thought about them, but had not acted on them. Reading the feedback comments is a private and 'low risk to ego' activity, and negative comments may easily be ignored or dismissed rather than addressed (Baumgardner et al., 1989).

The data from 10 of the EWF items was factored into three, more relevant sub-scales than those originally proposed. The results of this study suggest a small, positive, significant correlation between the LLE sub-scale and the OSCE scores. There was no significant association between the HLE sub-scale (which represented action) and OSCE scores as hypothesised in the SRL conceptual framework. There was a small, positive significant correlation between both the Reflection and Critical Reflection sub-scales and the OSCE scores, which was different to the results reported in Grant et al. (2006). It was decided to administer an updated questionnaire to a potentially larger sample of volunteer students from three cohorts to further explore the validity and reliability of the RQ and EWF scales.

Harrison et al.'s (2013) study at Keele Medical School in the UK reported how students at different performance levels made use of generic feedback (numerical breakdown of scores and skills) provided via a website. They noted 'Intriguingly,

the students who just passed the OSCE made least use of the feedback, yet they are at risk of failing future assessments and arguably have the most to gain from the feedback' (p742). Harrison et al. reported that they were not able to provide individualised examiner comments for their students and noted that a different pattern of engagement may have been recorded if these had been provided. One of the limitations of their study is that it is quantitative and recorded which students accessed generic and benchmark level feedback, but not if and how students acted on the information to improve their skills and knowledge. This study sought to provide data to begin to remediate the gap in understanding reported by Harrison et al.

7.6 Summary

This chapter outlines the development and administration of a scale which aimed to assess, via a hierarchical approach, students' engagement with the individual feedback comments they were provided with after their end of year clinical examination. Two sub-scales from the previously validated Reflection Questionnaire (RQ) were included to measure propensity to reflect on learning during professional courses of study, so that this important element of self-regulated learning could be assessed. The chapter discusses how the data from the Engagement with Feedback (EWF) questionnaire items were subject to factor analysis which resulted in better constructed sub-scales with stronger internal consistency values. Although analysis of the data from both sets of scales suggested some association with the level of scores from the clinical examination as was hypothesised, this was not as strong as was expected for the sub-scales which were of a theoretical higher order, Critical Reflection and Higher Level Engagement. In order to attempt to improve the EWF instrument, three of the EWF items were replaced and it was planned to administer an updated version of the EWF scale and repeat the administration of the two Reflection Questionnaire sub-scales to a larger number of students to further explore the two scales.

CHAPTER 8: STUDY 3 - DEVELOPMENT OF A SURVEY TO ASSESS ENGAGEMENT WITH FEEDBACK PHASE 2

8.1 Introduction

This Chapter describes the next empirical study in this thesis, the Phase 2 administration of an updated self-completion survey instrument. In light of the findings of the Phase 1 study described in Chapter 7, the three negatively scored Engagement with Feedback (EWF) items were replaced with three positively worded statements (as shown in Chapter 4, Table 4.1) and the 20 item survey instrument was administered to fourth, fifth and graduand students as a Phase 2 study in order to explore the validity and reliability of the modified EWF sub-scales and the temporal reliability and validity of the same Reflection Questionnaire (RQ) items that were administered in Phase 1. The Phase 2 results for the RQ items are reported in Chapter 6.

8.1.1 Research questions

The research questions this study aims to address via data generated from the questionnaire are similar to those in the Phase 1 study, but relate to the modified set of 12 EWF items:

1. Do the proposed Phase 2 set of EWF items effectively assess student engagement with the feedback comments provided after their clinical examinations?
2. Is the level of student engagement with feedback as assessed by the modified set of EWF items associated with their performance level in their end of year clinical examinations?
3. Is the students' propensity to reflect on their learning in general associated with their performance level in their end of year clinical examinations?

8.2 Research design

8.2.1 Survey instruments

Engagement with Feedback (EWF) Sub-scales

As described in Chapter 7, 10 of the initial 12 items in the EWF scale factored into three sub-scales. The three negatively scored items were replaced, so the Phase

2 EWF instrument again had 12 items in three proposed sub-scales; 'Thinking for Pass/Fail' (TPF) included three items related to thinking about performance in terms of whether the student had passed or failed the clinical examination overall, before the feedback was released. The second sub-scale, named Lower Level Engagement (LLE) included three items relating to reading and thinking about each of the feedback comments. The third sub-scale, named Higher Level Engagement (HLE) included four items relating to thinking about action, planning action or having taken action. The scale was presented in a Likert format described in Chapter 4. In the Phase 2 version of the EWF scale, none of these items were reversed scored.

Reflection Questionnaire Sub-scales

The two RQ sub-scales used were 'Reflection' and 'Critical Reflection'. Each of these scales had four items presented in a Likert format described in Chapter 4. None of these items were reversed scored.

Questionnaire Design

The two sets of scales were administered as one questionnaire, divided into three sections. Section A included the RQ scale items. Section B related to thoughts and actions before the feedback was available. Section C related to thoughts and actions after accessing the feedback. The version of the questionnaire completed by the students did not include sub-scale headings. Table 8.1 shows the items under their sub-scale headings and with the number order they were presented in. The full survey is shown as Appendix 4.

Table 8.1: Phase 2 Questionnaire Sub-scale Items

Reflection

1. I sometimes question the way others do something and try to think of a better way.
3. I like to think over what I have been doing and consider alternative ways of doing it.
5. I often reflect on my actions to see whether I could have improved on what I did.
7. I often re-appraise my experience so I can learn from it and improve for my next performance.

Critical Reflection

2. During this course I discovered faults in what I had previously believed to be right.
4. As a result of this course I have changed the way I look at myself.
6. This course has challenged some of my firmly held ideas.
8. As a result of this course I have changed my normal way of doing things.

Pass/Fail Thinking

9. I thought about the knowledge and skills I was asked to demonstrate at each station.
11. I asked other students how they had answered questions to help me to work out how I had done on each station.
12. I thought to myself about how I had done on each station.

Lower Level Engagement

13. I looked at **all** of my individual feedback comments.
16. I thought about whether I agreed or disagreed with each of my individual feedback comments.
17 I read my individual feedback comments to find out about my knowledge and skills.
18. I read my individual feedback comments **as soon as** I was able to access them.

Higher Level Engagement

10. I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills.
14. I thought about how I could use my individual feedback comments to improve my knowledge or skills.
15. I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.
19. I took specific action to improve my knowledge or skills as suggested in my individual feedback comments.
20. It is important to act on my individual feedback comments to improve my knowledge and skills in this academic year.

8.2.2. Participant recruitment

Participants were a convenience cross sectional sample of incoming fourth and fifth year students and outgoing graduands on the Medics course at the University of Birmingham, UK in 2016. Table 8.2 below shows the methods by which participants were recruited. No financial or other incentives were offered to students to encourage them to complete the questionnaire.

All graduands (students who had passed their finals examinations at main sit) were emailed with consent information about the study and invited to participate via the on-line version of the questionnaire.

Y5 students present at a whole cohort lecture on 4 July 2016 were invited to complete a paper copy of the survey. As insufficient completed surveys were returned, they were also asked if they would complete the survey at two split cohort lectures. Y5 students are randomly divided into two groups for their clinical placements. Half the cohort starts the academic year with the Acutely Ill Patient (AIP) rotation, the other half undertake rotations in other subjects. AIP students are split into two sub-groups (A & B) for the purposes of attending a lecture programme. Therefore up to half of the year cohort was given a follow up invitation to complete the questionnaire if they had not already done so at the whole cohort lecture.

Y4 students present at a whole cohort lecture on 20th June 2016 were invited to complete a paper copy of the survey. As insufficient completed questions were returned, the total cohort was emailed invited to complete an on-line version of the questionnaire. The total numbers of students present at the Y4 or Y5 lectures were not known.

Table 8.2: Study Participant Recruitment Methods

Cohort (N)	Invited to complete paper-based questionnaire	Email invitation to complete on-line questionnaire
Y4 2016-17 (349)	Cohort lecture 20-6-2016	Open from 18-7 to 5-8-2016
Y5 2016-17 (347)	Cohort lecture 4-7-2016 AIP rotation lecture group A on 7-9-2016 and group B on 21-9-2016	
Graduands 2015-16 (347)		Open from 6-6 to 24-6-2016

8.3 Results

8.3.1 Participants

Table 8.3: Dates and number of questionnaires completed

Cohort	Paper-based questionnaire	On-line questionnaire	Total N Respondents
Y4 2016-17	Cohort plenary lecture 20-6-2016 N= 69	Open from 18-7 to 5-8-2016 N= 32	101
Y5 2016-17	Cohort plenary lecture 4-7-2016 N= 50 Group A AIP rotation lecture 7-9-2016 N= 20 Group B AIP rotation lecture 21-9-2016 N= 11		81
Graduands 2015-16		Open from 6-6 to 24-6-2016 N=51	51

A total of 233 (101 Y4, 81 Y5 and 51 Graduand) students completed the questionnaire out of a possible total cohort of about 1050. Although the respondents were drawn from a convenience rather than a random sample, the chi square goodness of fit results presented in Table 8.4 below show that the demographic characteristics of the convenience sample were not significantly different to a random sample and so for these characteristics, the sample is broadly representative of the total population of students.

Table 8.4: Y4, Y5 & Graduand Respondent Demographic Data Compared to Total Cohort

	Responders N	Total Population N	Resps as % of total pop	χ^2 For Goodness Of Fit
Course				
Main	200	908	22.0	No significant difference in the proportion of Main and GEC/Dentists respondents compared to the total cohort population, χ^2 (1, n=229) =.023, p=.88
GEC/Dentists	29	135	21.5	
Missing	4			
Domicile				
Home/EEC	221	969	22.8	No significant difference in the proportion of Home/EEC and International respondents compared to the total cohort population, χ^2 (1, n=232) =1.818, p=.178
International	11	74	14.9	
Missing	1			
Gender				
Female	150	651	23.0	No significant difference in the proportion of Male and Female respondents compared to the total cohort population, χ^2 (1, n=233) =.559, p=.455
Male	83	392	21.2	
Age bands				
21 or less	51	167	30.5	No significant difference in the proportion of respondents in each of 4 age bands compared to the total cohort population, χ^2 (3, n=233) =7.306, p=.063
22-25	165	777	21.2	
26-30	16	89	18.0	
31+	1	10	10.0	

NOTE: GEC = Graduate Entry Course, graduate entry students undertake a separate first year of study and then join the main cohort in year 3. Dentists have a primary qualification in dentistry and directly enter the main course in year 3.

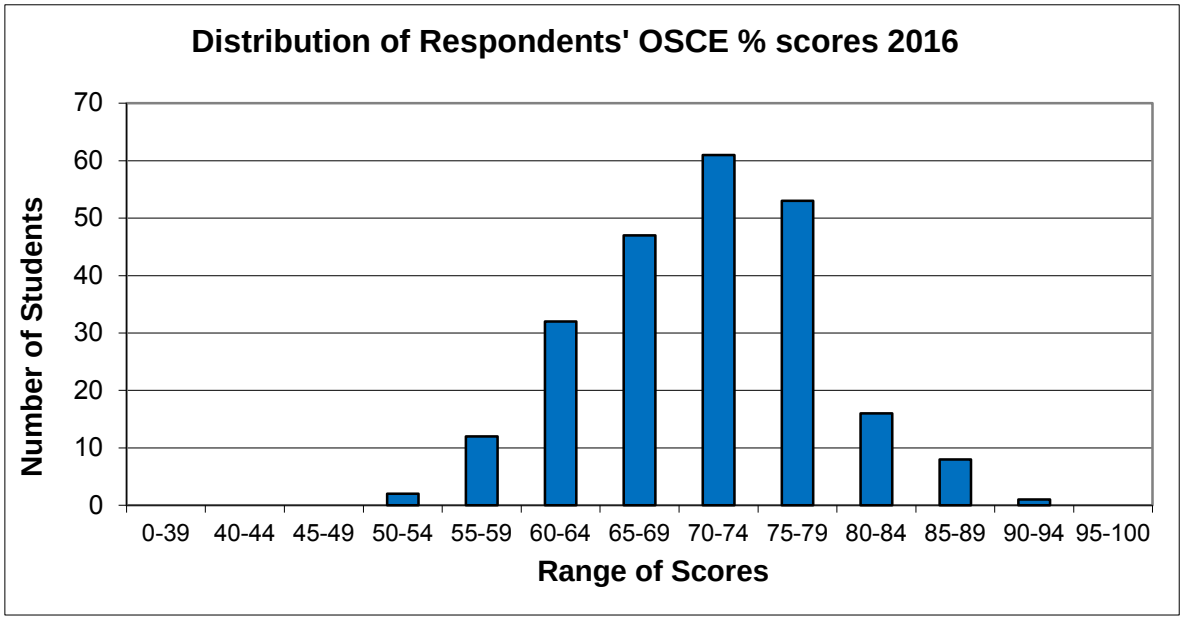
Table 8.5: Breakdown of Respondents Demographic Characteristics by Year

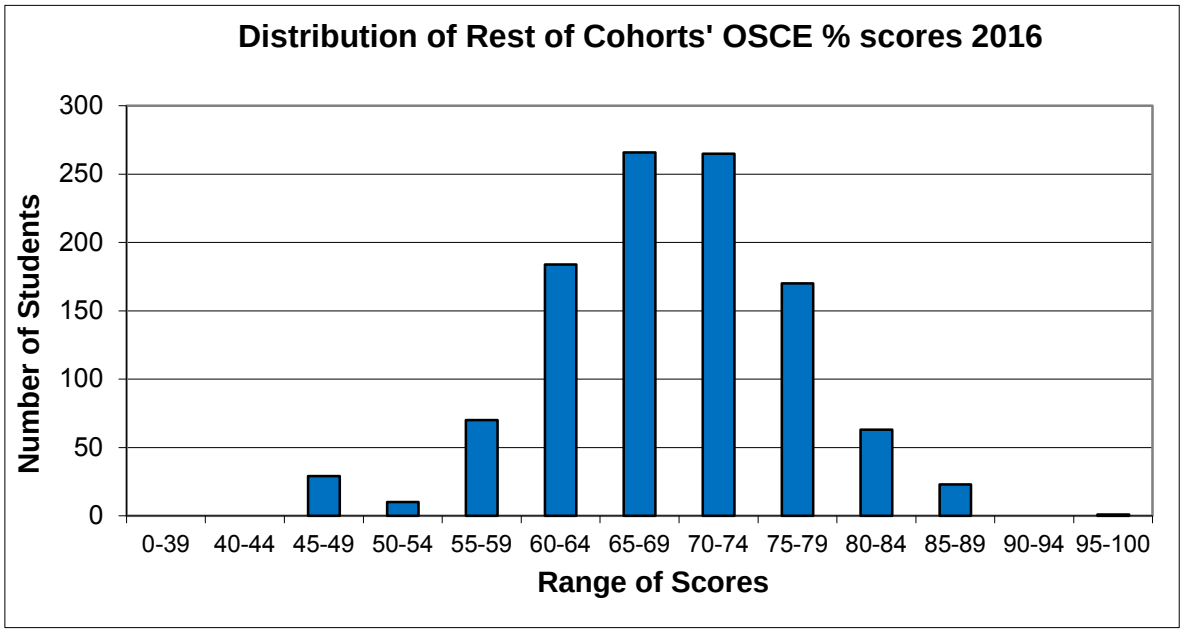
	Responders N	Total Population N	Resps as % of total pop	Responders N	Total Population N	Resps as % of total pop	Responders N	Total Population N	Resps as % of total pop
	Y4			**Y5**			**Granuands**		
Course									
Main	82	297	27.6	71	307	23.1	47	304	15.5
GEC/Dentists	16	52	30.8	9	40	22.5	4	43	9.3
Missing	3			1					
Domicile									
Home/EEC	93	323	28.8	79	329	24.0	49	317	15.5
International	7	26	26.9	2	18	11.1	2	30	6.7
Missing	1								
Gender							28	218	12.8
Female	71	218	32.6	51	215	23.7	23	129	17.8
Male	30	131	22.9	30	132	22.7			
Age bands									
21 or less	48	157	30.6	3	10	30.0	0	0	
22-25	47	172	27.3	72	305	23.6	46	300	15.3
26-30	6	17	35.3	6	30	20.0	4	42	9.5
31+	0	3	0.0	0	2	0.0	1	5	20.0

The distribution of scores achieved in the end of year clinical examination (OSCE) by the convenience sample appears to be different from those of the rest of the cohort, as shown in Figure 8.1 below. The convenience sample does not include any scores from the 45-49 range. These are first sit fail scores and although respondents were asked to provide the score for their first sit of the OSCE, previously failing students may have provided their supplementary sit score, or may not have volunteered to respond to the survey. There appear to be more respondents in the 75% and above bands, and less in the 55-74% bands. Only one respondent omitted to provide their OSCE score.

An independent-samples t-test was conducted using the guidelines in Pallant (2013, p250) to compare the OSCE scores for respondents and the rest of the cohort. There was a significant difference in scores for respondents (M=71.97, SD= 7.3) and the rest of the cohort (M=68.82, SD=7.8; t (386.5) = -5.715, p=<0.001, two-tailed). The magnitude of differences in the means (mean difference= -3.152, 95% CI: -4.236 to -2.068) was small (eta squared = 0.03).

Figure 8. 1: Distribution of OSCE % Scores for the Sample and the Rest of the Cohort





8.3.2 Data collation

The paper version of the questionnaire was designed to be optical mark recognition (OMR) machine read. Respondents were asked to fill in small circles to indicate their answers. The students are familiar with this process as all the multiple-choice examination answer papers use this layout. The completed questionnaires were scanned using Remark Office OMR (version 8) software, and a document scanner (Fujitsu fi-5530-C2) to extract the data.

The Bristol Online Survey tool was used to create the electronic version of the questionnaire. Respondents are directed to indicate their answers by clicking on 'radio' buttons. Respondents cannot proceed with later questions unless they complete an answer. The University of Birmingham has an account with this secure service provided by the University of Bristol. Completed survey data was downloaded from this site via Excel spreadsheets.

Students were given the option to provide their University registration number on the paper and electronic versions of the survey. If they did this, the Data Manager for the College of Medical and Dental Sciences at the University of Birmingham used this registration number to link the questionnaire Likert scale scores to the relevant demographic and OSCE score data for each respondent. A unique study participant ID code was assigned and the registration number deleted before passing the anonymised data to me.

The resulting data were initially imported into a Microsoft Excel 2010 spreadsheet for coding of the variables into data suitable for analysis in IBM SPSS Statistics 22 and for error checking. As suggested by Pallant (2013, p131), any cases with missing data were included in any of the analyses for which they provided the necessary information, but excluded where they are missing the data required for the specific analysis. This was achieved by the use of the 'exclude cases pairwise' option in SPSS 22.

For most of the analyses the results from the Y4 and Y5 students and graduands were combined and treated as one set of respondents.

8.3.3 Individual Item Statistics

The full range of scores was not selected for eight of the 20 questions. The minimum mean score was 3.35 and the maximum was 4.80 out of five (Table 8.6).

Table 8.6 Individual item statistics

Question No	Range	Mean score	Std. Deviation	N Respondents
1	1-5	4.08	.790	231
2	2-5	4.18	.826	233
3	1-5	4.15	.819	233
4	1-5	4.16	.947	232
5	2-5	4.29	.809	233
6	1-5	3.50	1.143	232
7	1-5	4.03	.861	233
8	1-5	3.76	.924	233
9	1-5	4.42	.758	232
10	1-5	3.59	1.362	233
11	1-5	3.82	1.348	233
12	2-5	4.72	.606	233
13	2-5	4.80	.594	220
14	2-5	4.26	.813	220
15	2-5	3.90	.972	220
16	1-5	4.53	.802	220
17	2-5	4.57	.641	220
18	1-5	4.16	1.288	220
19	1-5	3.35	1.059	220
20	2-5	4.40	.718	220

8.3.4 Initial sub-scale statistics

As expected, there were a number of significant correlations between the items which comprise the two RQ sub-scales (Reflection items 1, 3, 5, 7 and Critical Reflection items 2, 4, 6, 8). With regards to the EWF items, again there were a number of significant correlations between items on the same hypothesised sub-scales and between items on other sub-scales Thinking for Pass/Fail items 9, 11 and 12; Lower Level Engagement items 13, 16, 17 and 18; Higher Level Engagement items 10, 14, 15, 19 and 20) and between items on other sub-scales (Table 8.7).

Table 8.7: Pearson's Product-Moment Correlation Coefficient for the 20 RQ and EWF Items

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	_																			
2	**.191****	_																		
3	**.316****	.106	_																	
4	.106	**.180****	**.232****	_																
5	**.338****	.083	**.396****	**.290****	_															
6	**.158***	**.228****	**.186****	**.431****	**.270****	_														
7	**.277****	.103	**.373****	**.149***	**.534****	**.211****	_													
8	**.158***	**.231****	.115	**.419****	**.132***	**.445****	.116	_												
9	.032	.085	**.172****	.065	**.181****	.015	**.213****	.066	_											
10	.068	.089	.106	.018	.116	**.149***	**.215****	.032	**.275****	_										
11	-.096	.009	-.023	-.032	-.021	.061	-.033	.074	.090	.082	_									
12	.085	.085	.051	.019	**.167***	.036	.121	-.043	**.319****	**.129***	**.129***	_								
13	.045	-.073	-.050	.018	.122	.005	.011	.018	.104	-.037	.031	**.289****	_							
14	.080	.023	**.167***	.098	**.136***	.032	**.313****	.050	**.229****	**.199****	-.023	**.208****	**.223****	_						
15	.022	.027	**.210****	**.142***	.053	.044	.128	**.141***	**.251****	**.192****	.050	.075	.109	**.431****	_					
16	.133	.073	.059	.024	**.156***	.097	.122	.029	**.272****	.108	.058	**.286****	**.301****	**.141***	**.177****	_				
17	.088	**.163***	.123	.065	**.200****	.092	**.141***	**.189****	**.212****	**.141***	.056	**.154***	**.192****	**.246****	**.315****	**.227****	_			
18	.023	-.052	-.001	-.055	.003	-.042	.009	.075	.060	**.148***	**.187****	.081	**.263****	.095	.100	**.143***	**.144***	_		
19	.071	-.076	.132	-.028	.118	.060	**.169***	.105	**.234****	**.362****	.111	.064	.001	**.281****	**.449****	**.159***	**.295****	**.184****	_	
20	.071	**.142***	.097	.004	**.176****	.071	**.200****	**.198****	**.170***	**.194****	-.014	.071	**.148***	**.340****	**.402****	.045	**.421****	.123	**.410****	_

**Correlation is significant at the 0.01 level (2-tailed)

*Correlation is significant at the 0.05 level (2-tailed)

As expected, as shown in Table 8.8 below, there were a number of small to medium, positive significant correlations between sub-scales on the same scale, as both EWF and RQ sub-scales are assessing hierarchical, constructs, which are not mutually exclusive. In line with the theoretical framework, Lower Level Engagement (LLE) and Higher Level Engagement (HLE) correlated with Reflection and Thinking for Pass/Fail (TPF) did not. Only the LLE sub-scale indicated a small, positive correlation with OSCE scores. None of the other sub-scales were significantly associated. The results of Phase 1 indicated that the Reflection and Critical reflection sub-scales significantly correlated with OSCE scores, as well as the LLE.

Table 8.8: Pearson's Product-Moment Correlation Coefficient for RQ & Initial EWF Scales and Sub-scales with OSCE Scores

Scales	Reflection	Critical Reflection	Think P/F	LLE	HLE	RQ	EWF	OSCE Score
Reflection	-							
Critical Reflection	**.323****	-						
Think P/F	.088	.073	-					
LLE	**.136***	.063	**.297****	-				
HLE	**.238****	.125	**.203****	**.333****	-			
RQ	(.785**)	(.840**)	.097	.117	**.216****	-		
EWF	**.239****	**.131***	(.610**)	(.702**)	(.796**)	**.221****	-	
OSCE Score	.095	-.030	.018	**.176****	.005	.034	.084	-

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Note: Brackets highlight self-correlation scores of scales with their sub-scales

Cronbach's alpha score was calculated to assess the reliability (internal consistency) of the scales and sub-scales (Table 8.9). The values for two of the EWF sub-scales (TPF and LLE) were lower than the normally considered acceptable cut off of 0.7 (Pallant 2013, p104). Cronbach's alpha is usually positively influenced by the number of items included in the calculation and the five item HLE sub-scale reaches the minimum cut-off score. The results for the two Reflection sub-scales considered as one scale called RQ and the EWF items as

one scale are shown to explore if this might be a fruitful approach. However, the Cronbach's score for each scale does not increase substantially considering the number of items included because of the fairly low correlations between the sub-scales.

Table 8.9: Estimates of Reliability (internal consistency) and descriptive statistics for Scales and Sub Scales

Scale	Cronbach's Alpha	Mean	SD	Range
Initial Engagement with Feedback (EWF) Sub scales:				
Thinking for Pass/Fail (Q9, 11, 12)	.396	4.32	0.62	3.00
Lower level engagement with FB (Q13, 16, 17, 18)	.518	4.51	0.54	2.00
Higher level engagement with FB (Q10, 14, 15, 19, 20)	.707	4.87	0.73	4.00
12 Engagement with Feedback items	.727	4.21	0.45	2.00
Reflection Questionnaire				
Reflection (Q1, 3, 5, 7)	.704	4.13	0.60	3.50
Critical Reflection (Q2, 4, 6, 8)	.662	3.90	0.68	3.00
8 RQ items	.715	4.02	0.52	2.50

8.3.5 Factor Analysis

Assessment of the suitability of the questionnaire data for factor analysis

As per the processes outlined in the Chapter 7, preliminary analyses were conducted to ensure there was no violation of the assumptions of normality, linearity and scores per each question are evenly spread (homoscedasticity). Analysis of results for individual EWF items showed that the full range of scores was used for six out of the 12 items, with the lowest mean score 3.35 and the highest 4.80 out of a maximum possible score of 5. (Table 8.6 above) Therefore scores per question are not normally distributed. Pallant (2013, p59) quotes Tabachnick and Fidell's belief that 'with reasonably large sample sizes, skewness will not make a substantive difference' (2013, p80). As discussed in Chapter 7, the ratio of participants to items was checked. The sample maximum in this study was 233 and because some questions were not answered by up to 11 respondents, the sample minimum was 222. As there are 12 questions, there are

a minimum of 18 respondents per question and this is a satisfactory number to proceed with exploratory factor analysis.

Principal Components Analysis (PCA)

A method of Exploratory Factor Analysis (PCA) was conducted on this data set in order to evaluate, and begin to potentially validate the factors that were constructed based on the findings from the Phase 1 study. If the results of the initial PCA indicated items which loaded on an unexpected factor compared to the hypothesised factors or cross-loaded on multiple factors, those items were deleted and the PCA repeated until a parsimonious solution was achieved which balanced maximising the amount of the cumulative variance from the factors with minimising the number of factors. As detailed below, the factor structure that best fitted this data set was not a 12 question three factor solution as initially hypothesised, but two factors from 10 questions. Items 11 and 18 were deleted because their communalties values were below .2 and internal consistency (reliability via Cronbach's alpha coefficient) was improved by excluding them. Other solutions were examined, (including a three factor scale from 10 or 12 questions) but they yielded solutions that were less conceptually meaningful.

An initial analysis obtained eigenvalues for each factor and the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity were conducted to confirm that the data set was appropriate for factor analysis. The KMO test was used to verify the sampling adequacy for the analysis, and the Bartlett's Test of Sphericity was used to determine if correlations between items were sufficiently large for EFA. Bartlett's Test of Sphericity should reach a statistical significance of less than .05 in order to conduct an EFA. The Bartlett test of sphericity was significant (χ^2= 390.056, d.f.= 45, p < 0.00) and the KMO test result was 0.771 which is above the recommended .6 cut off. Given these results, the correlation matrix was considered suitable for EFA.

Inspection of the scree plot showed the slope becoming more horizontal at the forth data point indicating that three factors were appropriate. However the three factor solution did not make sense conceptually with regards to the items that were factored together. Costello and Osborne advise that using the Keiser criterion

(retain all factors with eigenvalues greater than one) does not always yield the best results for a specific data set (2005, p1). Khan notes 'Perhaps the most effective method of deciding how many factors to retain is a parallel analysis' (2006, p692). A parallel analysis was conducted consisting of a comparison of the eigenvalues from this data set with the eigenvalues for a data set which was randomly generated using Monte Carlo software (Watkins, 2000). This analysis suggests that a two factor solution was appropriate as shown in Table 8.10 below.

Table 8.10: Results of Monte Carlo PCA for parallel analysis random data generator

Component number	Eigenvalue from PCA	Criterion value from parallel analysis	Decision
1	3.034	1.3230	Accept
2	1.462	1.2262	Accept
3	1.108	1.1527	Reject

An oblique method of rotation, Direct Oblimin was performed to examine the dimensionality of the two new sub scales. This method was chosen because it was hypothesised that the items in the EWF sub- scales were not mutually exclusive, i.e. it is logical that a student could give positive answers to lower level engagement questions and higher level engagement ones.

Communalities between the variables ranged from .261 to .581.The two factors, after rotation, explained 30.34% and 14.62% of the common variance, and cumulatively 44.96%. Factor loadings are shown in Table 8.11 below and were all greater than the recommended cut off level of 0.30.

Table 8.11: Pattern and Structure Matrix for PCA with Direct Oblimin Rotation of Two Factor Solution of 10 Engagement With Feedback Items

	Pattern Matrix			Structure Matrix			Communalities	
		Component			Component			
		1	2		1	2		
L	Q9	.315	**.422**	Q9	.428	**.506**	Q9	.348
H	Q10	**.520**	-.039	Q10	**.509**	.099	Q10	.261
L	Q12	-.046	**.742**	Q12	.151	**.730**	Q12	.535
L	Q13	-.078	**.704**	Q13	.109	**.683**	Q13	.472
H	Q14	**.544**	.206	Q14	**.598**	.350	Q14	.397
H	Q15	**.734**	-.007	Q15	**.732**	.188	Q15	.536
L	Q16	.031	**.683**	Q16	.213	**.692**	Q16	.479
H	Q17	**.523**	.201	Q17	**.577**	.340	Q17	.370
H	Q19	**.788**	-.145	Q19	**.749**	.064	Q19	.581
H	Q20	**.736**	-.083	Q20	**.714**	.113	Q20	.516

Notes: Scale name codes are included next to question numbers: L= Lower level engagement, H= Higher level engagement. Major loadings for each item are in bold

As shown in Table 8.12 below, the estimates of internal reliability (Cronbach's Alpha) indicated that HLE was above the normally accepted minimum cut-off point of .70 and LLE was below it.

Table 8.12: Estimates of Reliability (internal consistency) and descriptive statistics for EWF Sub Scales After Factor Analysis:

Scale	Cronbach's Alpha	Mean	SD	Range
Lower level engagement with FB (Q9, 12, 13, 16)	.590	4.61	0.46	2.5
Higher level engagement with FB (Q10, 14, 15, 17, 19, 20)	.731	3.98	0.68	4

The resulting two sub-scales broadly kept together the items from the initially proposed sub-scales. The first sub-scale named Lower Level Engagement (LLE) retained two of the originally proposed items and two which were originally in the Thinking for Pass/Fail sub-scale. The second sub-scale, named Higher Level Engagement (HLE) included five of the original HLE items plus Q 17 (I read my individual feedback comments to find out about my knowledge and skills).
Two Phase 1 items were deleted, Q11 and 18. All three of the new Phase 2 items were factored.

Correlation analysis suggests small but positive, significant correlations between most of the individual EWF items. Where there was no significant correlation, the relevant item belonged in the other sub-scale, except for Q9 with Q13 which are both LLE items. (Table 8.13)

Table 8.13: Pearson's Product-Moment Correlation Coefficient for 10 Engagement With Feedback items

Items	Q9L	12L	13L	16L	10H	14H	15H	17H	19H	20H
Q9L	-									
12L	**.319****	-								
13L	.104	**.289****	-							
16L	**.272****	**.286****	**.301****	-						
10H	**.275****	**.129***	-.037	.108	-					
14H	**.229****	**.208****	**.223****	**.141***	**.199****	-				
15H	**.251****	.075	.109	**.177****	**.192****	**.431****	-			
17H	**.212****	**.154***	**.192****	**.227****	**.141***	**.246****	**.315****	-		
19H	**.234****	.064	.001	**.159***	**.362****	**.281****	**.449****	**.295****	**-**	
20H	**.170***	.071	**.148***	.045	**.194****	**.340****	**.402****	**.421****	**.410****	-

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

As shown in Table 8.14 below, correlation analysis of the scales and factored EWF sub-scales suggests positive, small but significant correlations between the two sub-scales from the 10 EWF items, which was expected as these are hypothesised to be measuring related concepts. Both LLE and HLE correlated significantly with Reflection. HLE also significantly correlated with Critical reflection but LLE did not. This finding is in line with the hypothesised hierarchical relationship between the constructs represented by the two sub-scales in each of the scales. However, none of the scales or sub-scales significantly correlated with the OSCE scores.

Table 8.14: Pearson's Product-Moment Correlation Coefficient for Scales and Factored Sub-scales with OSCE Scores

	Reflection	Critical Reflection	LLE	HLE	RQ	EWF	OSCE Score
Reflection	-						
Critical Reflection	**.323****	-					
LLE	**.218****	.080	-				
HLE	**.245****	**.137***	**.271****	-			
RQ	(.785**)	(.840**)	**.177****	**.228****	-		
EWF 10Q	**.300****	**.148***	(.646**)	(.888**)	**.267****	-	
OSCE Score	.095	-.030	.128	.005	.034	.067	-

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Note: Brackets highlight self-correlation scores of scales with their sub-scales

The correlations overall indicated weaker associations between the factored sub-scales and the OSCE scores than in Phase 1, so further statistical analysis of the data was undertaken to investigate if any important variables in the sample had an unexpected influence on the results. Independent Samples t tests were conducted to ascertain if there were any differences between the mean scores from two of the cohorts, Y4 and Graduands (the two most different in terms of seniority) for the Reflection, Critical reflection, LLE and HLE sub-scales and OSCE scores. The results suggested that there were no differences between Y4 and Graduand respondents. The same comparisons were undertaken with gender as the variable. As shown in Table 8.15 below, the only significant difference was between the scores for the Critical reflection scale, with the male average score being higher(M = 4.02, SD = 0.68) than females (M=3.83, SD= 0.67; t (231) =-1.98, p =.05, two-tailed, d = -0.26). The same comparisons were also undertaken with Course (Main or Graduate Entry) as the variable. As shown in Table 8.16 below, the only significant difference was between the scores for the Reflection sub-scale, with the Graduate Entry Course (GEC) average score being higher (M = 4.40, SD = 0.54) than the Main course (M = 4.10, SD = 0.6; t (227) = -2.51, p = 0.01, two tailed, d = -0.33). About 36% of the respondents overall were male, and 13% of respondents (male and female) were on the GEC. Only one sub-scale result was significantly influenced by each group, so this probably did not have an effect on the results overall.

Table 8.15: Independent-samples t-test Comparing Female and Male Respondents

	Gender	Mean	SD	t	df	Cohen's d	Sig (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Reflection	F	4.13	0.61	-0.20	231	-0.03	.84	-0.02	-0.18	0.14
	M	4.15	0.57							
Critical Reflection	F	3.83	0.67	-1.98	231	**-0.26**	**.05**	-0.18	-0.37	0.00
	M	4.02	0.68							
LLE	F	4.61	0.49	-0.36	231	-0.05	.72	-0.02	-0.15	0.10
	M	4.63	0.42							
HLE	F	4.03	0.68	1.59	231	0.21	.11	0.15	-0.04	0.33
	M	3.88	0.69							
OSCE Scores	F	71.92	7.65	-0.15	230	-0.02	.88	-0.15	-2.14	1.83
	M	72.07	6.80							

Table 8.16: Independent-samples t-test Comparing Main and Graduate Entry Course Respondents

	Course	Mean	SD	t	df	Cohen's d	Sig (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Reflection	Main	4.10	.60	-2.51	227	**-0.33**	**0.01**	-0.30	-0.53	-0.06
	GEC	4.40	.54							
Critical Reflection	Main	3.89	.67	-0.17	227	-0.02	0.87	-0.02	-0.29	0.25
	GEC	3.91	.81							
LLE	Main	4.60	.47	-1.21	227	-0.16	0.23	-0.11	-0.29	0.07
	GEC	4.72	.43							
HLE	Main	3.96	.70	-1.19	227	-0.16	0.23	-0.16	-0.43	0.11
	GEC	4.12	.59							
OSCE Scores	Main	71.77	7.35	-1.38	226	-0.18	0.17	-2.02	-4.90	0.86
	GEC	73.79	6.50							

To further explore potential reasons why the results for the sub-scale scores and the OSCE scores were not as strongly associated as hypothesised, I considered if the combined average results of the clinical examination and the multiple choice question (MCQ) examination might have been a more appropriate measure of students' academic performance to use. This could not be calculated for respondents in this study, so correlations of the scores for both assessments for the whole cohorts were calculated and there is a medium sized, positive,

significant association between these two assessments for the three cohorts as a whole (Table 8.17 below).

Table 8.17: Pearson's Product-Moment Correlation Coefficient for Mean OSCE and MCQ Scores in 2016

		Mean	SD	N	r
Y3-5	OSCE	69.63	7.55	1081	.475**
	MCQ	71.49	7.98	1081	
Y3	OSCE	70.45	7.84	369	.491**
	MCQ	68.81	8.57	369	
Y4	OSCE	70.25	7.85	353	.567**
	MCQ	72.99	7.61	353	
Y5	OSCE	68.16	6.70	359	.479**
	MCQ	72.76	6.94	359	

** Correlation is significant at the 0.01 level (2-tailed)

As the individual item statistics (Table 8.17 above) revealed high mean scores for most of the items, the percentage of respondents who answered 'agree' or strongly agree' for each item was calculated in order to facilitate interpreting the pattern of results per item. The results are presented as Table 8.18 below with items in their sub-scales. This implies that there was a very high level of agreement with all the statements included in the Reflection sub-scale. There was less agreement for two of the items on the Critical reflection sub-scale. Similarly, there was high agreement with all items on the LLE sub-scale, but much less agreement with the items which related to having taken action on the HLE sub-scale.

Table 8.18: Percentage of Phase 2 Responders Who Agreed or Strongly Agreed With the Item Statements with items shown in relevant sub-scales

Sub-Scale	N	Question	% of respondents who agreed or strongly agreed
Kember Reflection	1	I sometimes question the way others do something and try to think of a better way.	90
	3	I like to think over what I have been doing and consider alternative ways of doing it.	89
	5	I often reflect on my actions to see whether I could have improved on what I did.	91
	7	I often re-appraise my experience so I can learn from it and improve for my next performance.	85
Kember Critical Reflection	2	During this course I discovered faults in what I had previously believed to be right.	86
	4	As a result of this course I have changed the way I look at myself.	85
	6	This course has challenged some of my firmly held ideas.	59
	8	As a result of this course I have changed my normal way of doing things.	73
Lower level EWF	9	I thought about the knowledge and skills I was asked to demonstrate at each station. [Before FB was available]	95
	12	I thought to myself about how I had done on each station. [Before FB was available]	97
	13	I looked at all of my individual feedback comments.	96
	16	I thought about whether I agreed or disagreed with each of my individual feedback comments.	93
Higher level EWF	10	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills. [Before FB was available]	65
	14	I thought about how I could use my individual feedback comments to improve my knowledge or skills.	90
	15	I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.	73
	17	I read my individual feedback comments to find out about my knowledge and skills.	95
	19	I took specific action to improve my knowledge or skills as suggested in my individual feedback comments.	54
	20	It is important to act on my individual feedback comments to improve my knowledge and skills in this academic year.	91
Deleted after FA	11	I asked other students how they had answered questions to help me to work out how I had done on each station. [Before FB was available]	76
	18	I read my individual feedback comments as soon as I was able to access them.	78

8.3.6 Answers to Research Questions

1. Do the proposed Phase 2 set of EWF items effectively assess student engagement with the feedback comments provided after their clinical examinations?

With regard to the first research question, the data from this study shows that the originally proposed three sub-scales were not as effective at assessing a hierarchical approach to engagement with feedback compared to the two sub-scales produced after factor analysis.

2. Is the level of student engagement with feedback as assessed by the modified set of EWF items associated with their performance level in their end of year clinical examinations?

With regard to the second research question, neither of the scores for the EWF sub-scales (LLE or HLE) was significantly associated with performance level in the end of year clinical examinations.

3. Is the students' propensity to reflect on their learning in general associated with their performance level in their end of year clinical examinations?

With regard to the third research question, neither of the scores for the RQ sub-scales (Reflection and Critical reflection) was associated with performance level in the end of year clinical examinations.

8.4 Discussion

Respondents in this study were given the opportunity to provide their University registration number so that their demographic data and OSCE score could be linked to their questionnaire answers. This was undertaken by the College of Medical and Dental Sciences Data Manager so that the responses were anonymous to the author. This ensured that this element of the data had no intentional or unintentional errors. The majority of students did give their registration number (Y4 87%, Y5 95%, Graduands 88%). This adds to the validity of responses in this study.

In response to free text comments on the Phase 1 study which said that respondents had not read the feedback because they had intercalated,

respondents in Phase 2 were additionally asked if they had intercalated or taken leave of absence in the previous year so this could be linked to whether or not they indicated that they had not read the feedback. Five out of 101 fourth year students, one of whom reported that they had intercalated, and eight out of 81 fifth year students, again one of whom reported that they had intercalated, reported that they had not read the feedback. None of the graduands reported that they had not read the feedback.

In the Phase 1 study, the reverse scored questions (Q9, 17, and 20) included a rationale for why students might not do or think that (e.g. Q9 'I did not think about how I had done on each station because it was too late to change anything'). Clark and Watson (1995, p312) caution against this complex or double-barrelled approach as responders might have had another reason for not thinking about the comments. Also Q17 and Q20 did not perform well in the psychometric analyses of the Phase 1 data and were removed from the results for the factored sub-scales. Therefore these three questions were changed for this study (shown in Chapter 4, Table 4.1).

Based on the findings from the Phase 1 data set, a three sub scale instrument was designed for use in Phase 2 but the results of the factor analysis in this study suggested that a two factor solution was more appropriate (Table 8.11 above). The resulting factors merged two of the three previous TPF scale items into the LLE scale and Q17 moved from LLE to HLE. All other questions remained in their original Engagement with Feedback sub-scale except for Q11 and 18 which were deleted from the instrument to improve the amount of shared variance and the internal consistency of the two new factors.

Table 8.14 above shows that although none of the sub-scales are associated with the OSCE scores, there are small, positive, significant associations between the Reflection sub-scale and both the EWF sub-scales, LLE and HLE. There is also a significant association between Critical reflection and HLE. Although causality cannot be determined, this correlation is interesting and intimates that the propensity to reflect is linked with propensity to engage with feedback. Also interesting is the association between the two sub-scales which are theorised to

be higher constructs in both scales, i.e. CR and HLE which provides some evidence of construct validity.

Table 8.18 above shows the percentage of respondents per item who agreed or strongly agreed with the item. This reveals very high levels of agreement for most of the items on the two EWF sub-scales, so this is probably why there is no association with the OSCE scores which are more normally distributed. 91% of students agreed or strongly agreed that it was important to act on the feedback, yet only 54% of them agreed or strongly agreed that they had taken specific action. 73% of respondents agreed or strongly agreed that they planned to take action during the academic year. More students reported that they acted to improve their knowledge or skills before receiving the feedback, than afterwards (Q10 65% and Q19 54%). As noted in Chapter 5, respondents may have undertaken action before receiving the feedback as an attempt to work out if they had passed or failed the examination rather than being intrinsically motivated to improve knowledge or skills.

Shute suggests that for students to take action resulting from feedback they need 'motive, opportunity and means' (2008, p175). In 2016, the incoming Year 4 started with a plenary day on 20th June and Y5 on 4th July. Table 8.3 above shows that 65% (119 out of 182) of Y4 and Y5 students answered the questionnaire at their plenary session, so it could be contended that they would not have had time to use their clinical placements to update their skills or knowledge. The other students completed the questionnaire some weeks later so would have had opportunity and means.

Before they commenced the current Y4, students were undertaking academic activities, not clinical placements, directly after their Y3 examinations, so they would only have been able to look up information, not improve clinical skills in a clinical environment. Directly after their Y4 examinations, the current Y5 went on elective placements. These may have been in the UK or overseas and may have been laboratory or clinically based, so these students may have had opportunities for action. After their final examinations, the graduand Y5 cohort undertook a four week long 'Assistantship' clinical rotation starting on 16th May 2016 and their

examination results and feedback were released on 24th May 2016, they had a period of time when they could have acted on the feedback before completing the on-line survey and the potential incentive to do this before starting their first post as junior doctors on 1 August 2016. Thus only some of the respondents may have had the 'motive, opportunity and the means' to act on the feedback.

With regard to the results for the RQ sub-scales, Table 8.18 also shows that the majority of respondents agreed with the RQ sub-scale items with between 59% to 91% agreeing or strongly agreeing with individual statements and an average per sub-scale of 89% for Reflection and 76% for Critical reflection, so that the majority respondents claim to be very reflective about their learning activities on the Medics' course.

Lucieer et al. assert that self-regulated learners are more effective learners who attain higher grades (2016, p589). However, after administering the 'Self-Regulation of Learning Self Report Scale' to 949 first and third year 'pre-clinical' medical students in Holland (which included a five item sub-scale on reflection), their results also suggested that the propensity to reflect was not related to academic performance. Their study confirmed that some of the variation in performance could be explained by the students' SRL skills but reported that a large part of the variance remained unexplained. In their Y1 cohort, reflection was found to be one of the skills related to grade point average (GPA) score. However, in their Y3 cohort, reflection was not related to academic performance but Y3 students had higher levels of reflection than Y1 students. This is similar to my findings in terms of there being no association between reflection and clinical examination scores, but I did not find any significant differences between the end of year clinical examination results for Y3 and Y5 students.

Limitations of this study

This study is predicated on the hypothesis that students with effective self-regulated learning skills would have a higher propensity to cognitively engage with and act on, their individual feedback comments and to reflect on their learning on their course of study. However it has been noted that students may be more or

less effective self-regulated learners depending on the context and self-regulated learning is not a uniform trait (Sandars 2013, p1162).

The RQ sub-scales aim to measure reflection on learning on the course and it was assumed that the propensity to reflect would also apply to reflection about a summative examination. However, it may be the case that students do not realise that the clinical examination is also a learning opportunity because it may be seen as simply a hurdle they have to pass to progress, rather than a dual opportunity to also reflect on and audit their knowledge and skills against the expected standards for the questions or tasks presented in the examination. It might have been more discriminating to use one of the RQ sub-scales which aimed to assess non-reflective thinking, such as 'Understanding' with the Reflection sub-scale in order to better distinguish between the types of thinking when attempting to assess any association with the scores from the clinical examination.

Tricio et al. (2015) administered the full RQ four sub-scale survey to 172 dental students in their third, fourth or fifth year of study and found a positive, significant correlation between the OSCE score and the Habitual action and Understanding sub-scales only. Similar to my findings, they found no association with the Reflection or Critical reflection sub-scales. They account for this by explaining that students are not expected to demonstrate Reflection or Critical reflection in competency examinations, but are given the opportunity to show that they have developed their 'understanding' and decision making skills by learning 'to restrict themselves to the relevant features and aspects of the presented situation'. They explain the correlation with the Habitual action sub-scale by the fact that OSCEs often include common clinical cases or expected questions that could be approached by competent students in a habitual way (2015, p119). I had not considered this important aspect when I decided to use the OSCE scores as the proxy measure for SRL abilities. The clinical examination is a performance based assessment, the format of which is daunting to some students (Hodges, 2003) and this may have had an adverse effect on the level of performance that might usually be predicted for some respondents.

The scores from the clinical examinations were used as a proxy measure for students' self-regulated learning ability as I hypothesised that effective self-regulated learners would score more highly in the OSCE. The association between the EWF sub-scales and the OSCE scores were examined as an aspect of validating the sub-scales, with the expectation that there would be a stronger association with the HLE sub-scale than the LLE sub-scale. However, for the reasons explained above, the OSCE scores may not have been an accurate proxy measure for self-regulation. It may have been better to ask students for their overall score from their end of year assessments for the Clinical Core module (which would include the result of the multiple choice question examination). As shown in Table 8.17 above, there is a medium strength association (r=.475) between the scores for the OSCE and Multiple Choice Question (MCQ) paper achieved by the relevant student total cohorts in 2016, so using the combined average scores of respondents might reveal a different association with the RQ or EWF scales. Another approach would have been to administer a questionnaire which measured self-regulated learning propensity. However, the concept of SRL includes many different elements to assess. The validated (for use with school aged children) Motivated Strategies for Learning Questionnaire (MSLQ) has 44 items (Pintrich and De Groot, 1990, Pintrich et al., 1993), so it would not have been feasible to administer a combined questionnaire of 64 items in the time available with students in this context.

It is possible that the EWF sub-scales do not effectively capture the intended latent constructs. The descriptive statistics for the individual items show that many items were not discriminating because the full range of scores was not selected. This may have been the influence of social desirability bias (Streiner et al. 2015, p106) causing students to over-estimate their self-reported engagement with feedback or reflection activities. One method to counteract this would be to use a 'positively packed' rating scale with more positive than negative scale descriptors (Brown et al. 2016, p613). Another reason why the instrument may not have effectively captured the intended constructs may be because there were too few items per sub-scale.

The first four EWF items (Q9-12) about thoughts or actions *before* receiving the feedback aimed to assess the level of self-assessment and feedback before the influence of the external feedback from examiners. This may have complicated the scale by introducing another, totally separate latent construct. The other eight questions sought to capture the difference between accessing and thinking about the feedback and the higher level engagement construct, acting (or planning to act) on the feedback in order to improve knowledge or skills.

The number of respondents in this study may also be an issue. Although there were 233 respondents, which were sufficient for the 'rules of thumb' required for factor analysis quoted by Pallant (2013, p190), Khan quotes Comrey and Lee's (1992) suggestion that '100 cases are poor, 200 are fair, 300 are good and 500 or more are very good' (2006, p700). Kahn believes that sample size decisions should not be driven by the number of variables being analysed but rather if at least three or four variables have high structure coefficients for each factor and communalities are high enough, i.e. .6 and above (2006, p700). As shown in Table 8.11 above, the loadings on each factor in the structure matrix ranged from .506 to .730 for LLE and .509 to .749 for HLE and communalities ranged from .261 to .581, so the Phase 2 data meets only one of these standards. Although the sample was representative of the whole cohort, it may simply have not been large enough to produce reliable results.

8.5 Conclusion

The Phase 2 study aimed to answer similar research questions as the Phase 1 study and had different findings in that there were no significant associations between the OSCE scores and any of the four sub-scales. The Phase 2 respondent sample was larger and so more likely to be reliable. However, as three of the questions had changed in the EWF sub-scales, this may also have had an influence on the results, although the RQ items were exactly the same. The reasons why the RQ scale and EWF scale were not associated with the OSCE scores as hypothesised may be because the clinical examination was not an effective proxy measure for SRL abilities or activities; the EWF scale was not a valid measure of engagement with feedback; or that the RQ was not valid measure of reflection. As shown in Chapter 6, the RQ is a reliable measure and

the findings from the Phase 1 and 2 studies add to the evidence from other studies. However, studies by Grant et al. (2006) and Lucieer et al. (2016) also found that propensity for reflection is not related to academic performance in assessments. As Ablard and Lipschultz reported after a study with high-achieving 7th grade children, 'The relation between achievement and self-regulated learning is more complex than originally believed' (1998, p94).

Considering the questionnaire items individually, rather than as sub-scales, the results in Table 8.18 are reassuring in that the majority of the respondents claim to have accessed and cognitively engaged with the feedback. About half of the students also claim to have acted on the feedback in some way. This lower score may be because students were not motivated to act or because of the timing of the administration of the survey. Based on the interview findings in Chapter 5, at least some of the feedback comments should have provided specific, actionable suggestions for improvement. The results in Table 8.18 for the RQ sub-scales also show that the majority of respondents agreed with these items. This answers my initial questions posed in Chapter 1; whether the performance level of medical students in their end of year clinical examinations is associated with their level of engagement with the individual feedback comments provided after the exam and/or their propensity to reflect on learning events generally. Although there is no significant association with scores from the OSCE, it would appear that the majority of respondents do have some level of engagement with the feedback comments and that the majority of respondents do reflect on their learning activities.

The psychometric indices of the new EWF scale were not particularly strong. Therefore further work needs to be done to improve these before the scale can be used as a generalisable method of assessing engagement with feedback from academic assessments. This may be achieved by the inclusion of additional items. The development of a valid and reliable instrument is a lengthy process requiring multiple studies across periods of time and across different undergraduate medical programmes or other healthcare professions settings. This study is an additional step in developing an instrument to assess engagement with

feedback from clinical assessments but further studies are required to develop the quality of this instrument in an iterative fashion.

8.6 Summary

In Phase 1, a new 12 item survey instrument to assess Engagement with Feedback (EWF) was evaluated for the psychometric properties of reliability and dimensionality. In Phase 2, three of the EWF questions used in Phase 1 were changed and the reliability and dimensionality of two slightly different sub-scales were evaluated. The Phase 2 study aimed to answer the same research questions as the Phase 1 study. The findings were that there was no significant association between the RQ and EWF sub-scale scores with the OSCE scores in Phase 2. The RQ sub-scale items were the same in both studies. The Phase 2 sample was larger and representative of the population, so was more likely to be accurate than the Phase 1 data. Reasons were explored why it may not have been appropriate to use the OSCE scores as a proxy measure for SRL ability. However, there was a small, statistically significant correlation between the EWF and RQ sub-scales, which reveals a link between propensity to reflect and propensity to engage with the feedback, although causality cannot be established.

Consideration of the individual items in both EWF sub-scales shows that the majority of the responders claim to have cognitively engaged with their feedback comments and agree that it is important to act on the feedback, but a smaller number claim to have taken action on their feedback in order to improve their skills or knowledge. This lower score may, in part, be due to the timing of the administration of the survey. Similarly, the majority of respondents agreed with the RQ sub-scales items. Further work needs to be done to improve the convergent and discriminant validity of the new EWF scale if it is to be a generalisable method of assessing engagement with feedback from academic assessments.

CHAPTER 9: CONCLUSION

9.1 Introduction

In this chapter I will provide a brief review of the findings from the three empirical studies presented in this thesis and sum up the original theoretical and empirical contributions to knowledge made by this thesis study. I will outline some suggestions for further studies that could be undertaken to build on these findings, in this context or elsewhere and make some recommendations for improving student engagement with their feedback in this context which are potentially generalisable to other learning contexts.

The focussed review of the literature presented in Chapter 2 shows that that the giving and receiving of feedback is an important element of learning. There are established theoretical frameworks about the sort of feedback which should be provided to improve understanding of the subject matter or metacognitive skills (Black and Wiliam 1998; Nicol and Macfarlane-Dick 2006; Hattie and Timperley 2007; Shute 2008; Wiliam 2011). In higher education much attention has been paid to aspects of giving and receiving written feedback on summative or formative, essay type assessments across a range of subject disciplines, from the viewpoints of both the tutors who provide the feedback and the student recipients (e.g. Higgins et al. 2002; Bailey and Garner 2010; Orsmond and Merry 2011; Hepplestone and Chikwa 2014).

However, with regard to practice based, 'live' summative or formative undergraduate clinical examinations, there is currently a paucity of similar research. This is surprising, given that OSCE type examinations are a ubiquitous method of assessment in undergraduate (and postgraduate) medical and other healthcare professions' education in the UK and the rest of the world (Boursicot et al., 2010) and that it has been widely reported how important it is for clinicians to develop skills in seeking and acting on feedback to improve their practice (e.g. Sargent et al. 2008; Mann et al. 2009; Sandars 2009; Pelgrim et al. 2013).
This lack of research into engagement with feedback after clinical examinations may be because feedback is not commonly provided after such assessments due

to logistical difficulties, (especially if paper mark sheet pro formas are used), or because faculty do not view summative assessments as also having a legitimate function in generating formative feedback to confirm how, at least minimal standards of competency were achieved, and to provide suggestions for improvement in future healthcare practice, (or for failing students, also to improve performance in future iterations of an examination). I was instrumental in ensuring that a method of delivering individual feedback comments to cohorts of 350-400 students was implemented, despite using paper mark sheets. As advised by Shute (2008), these feedback comments were descriptive about the current state of the student's performance and prescriptive in suggesting appropriate courses of action for improvement. What was not known before this thesis study was conducted was if, and how students at Birmingham Medical School engaged with this individualised feedback; if they accessed it, thought about it or acted on it; and if students at different levels of performance in the OSCE had different approaches. This led to the overarching research questions below:

Interview study

1. How do students at different levels of performance engage with their individual feedback comments?
2. Is the Self-Regulated Learning (SRL) conceptual framework appropriate for describing the cognitive and concrete actions relating to engagement with feedback?

Survey studies

1. Do the proposed Phase 1 and 2 Engagement with Feedback sub-scales effectively assess student engagement with the feedback comments provided after their clinical examinations?
2. Is the level of student engagement with feedback associated with their performance level in their end of year clinical examinations?
3. Is the students' propensity to reflect on their learning on their course associated with their performance level in their end of year clinical examinations?

9.2 Review of findings

Before discussing the findings of these three studies, I would like to remind readers about the nature of the data presented in this thesis. The information from participants is self-reported and so the resulting data is predicated on what students chose to reveal. I am aware that it is possible that students may not have faithfully reported their thoughts or actions in the interviews or the surveys. However, there was no plausible benefit to them not being truthful with me, apart from the usual propensity for people to describe themselves in the best light possible, and this may have led to an element of conscious or unconscious inflation of agreement with the survey items.

The first empirical study aimed to assess if the outline SRL conceptual framework devised to describe effective engagement with feedback did in fact accurately represent this. Interviews were conducted with eleven Y5 students who were stratified into three performance levels by their Y4 OSCE scores (four in the top, four in the borderline passing and three in the first sit fail groups). The students were asked to discuss if and how, they engaged with their individual feedback comments. The top performing students described effective engagement with the feedback which included reflecting on their performance and acting to remediate self-assessed deficits *before* the feedback was released as well as critically reflecting on their feedback comments and elucidating a rationale for acting on them or not. They understood that acting on the feedback was for their benefit in developing as clinicians. The resit and borderline passing students tended to attempt to self-assess whether they had passed or failed the examination before receiving the feedback, rather than reflecting on specific strengths and weaknesses in their knowledge and skills. The resit students had the external motivation to act on the feedback, but usually framed this as aiming to improve their performance in the resit examination only, not for future clinical practice. The borderline students tended not to act on the feedback. The fact that students at all three performance levels reported they understood the feedback and knew how to act on it, even though they may not have acted, was a promising start to investigating this issue.

The interview study enabled further details to be added to the outline SRL conceptual framework, such as student self-assessment before the feedback was available, and the process of reconciling internal and external feedback as described in Chapter 3. The final version of the conceptual framework did appear to be appropriate for describing the cognitive and concrete actions relating to engagement with feedback in a hierarchical way, and was used to inform the design of the Engagement with Feedback (EWF) survey items, which aimed to assess a larger sample of students' thoughts and actions with regard to the feedback. As reflection on performance in the clinical examination was held to be an important facet of SRL in the conceptual framework, two sub-scales from a previously validated survey instrument, the Reflection Questionnaire (Kember et al., 2000) were included with the EWF items.

180 Y4 and Y5 students responded to the Phase 1 survey with regard to their end of Y3 or Y4 clinical examinations. The results from a principal components analysis (PCA) suggested a three factor solution of ten items. The results of the Pearson correlation analysis indicated that there were small but statistically significant associations between the OSCE scores and the EWF sub-scale Low Level Engagement (LLE), as well as the two RQ sub-scales, Reflection and Critical Reflection. Three of the EWF items which had not performed well in the PCA were modified, and 233 Y4, Y5 and graduand students responded to the Phase 2 administration of the combined survey instrument. The results from the PCA suggested a two factor solution of ten items for the Phase 2 data. The results of the Pearson correlation analysis of the Phase 2 data (which included a third more responders than Phase 1) indicate that neither the respondents' propensity to reflect on their learning on this professional course nor their level of engagement with the feedback comments from the end of year clinical examinations was significantly associated with their scores in the clinical examination.

Both the Phase 1 and Phase 2 results suggest statistically significant associations between the EWF and the RQ scales and sub-scales, providing evidence that these two constructs are related as hypothesised in the SRL conceptual framework, but a causal relationship cannot be asserted from these results.

The results per item for the EWF scales in both Phase 1 and 2 demonstrated that the majority of respondents agreed or strongly agreed with the statements relating to cognitive engagement with the feedback, and had lower agreement with the items regarding action or thinking about action on the feedback. However, in Phase 2, the majority of students agreed with the new item statement that it was important to act on their feedback. This finding indicates a gap in the current level of engagement with feedback by many respondents between thinking about the feedback, knowing that it is important to act on it, but not actually doing this.

As noted above, the Phase 1 and 2 survey studies included the administration of two sub-scales, Reflection and Critical reflection from the previously validated Reflection Questionnaire (RQ). The results from the principal components analysis (PCA) for the Phase 1 RQ data were not as theoretically expected, with two items factoring onto the opposite sub-scale to the one expected. The items did factor as expected in Phase 2. With reference to the individual items in these two sub-scales, in Phase 1 and 2 the students reported high levels of agreement with the statements which indicates that they have a high propensity to reflect on their learning during the course, and so it should be possible to build on this by highlighting to students that they should also reflect on how they could act to improve skills and knowledge and subsequently, if they were successful in achieving this.

In a recent publication, Brown et al. (2016) note that the assumption that students who are effective self-regulated learners will make use of feedback has not been empirically tested, and the Phase 1 and 2 EWF surveys represent an attempt to do this. One weakness of the Phase 1 and 2 survey studies is that I assumed that effective self-regulated learners would score highly in the clinical examination, so these scores would be an appropriate proxy measure for self-regulated learning. I did not include sub-scales to independently assess any aspects of the students' SRL ability as Brown et al. did. Interestingly, they reported that they did not find a direct association between the scores from the SRL 'Monitoring' sub-scale and the grade point average (GPA) scores of their 278 university student respondents. However, they did find a positive association between one of the five sub-scales in

their ‘Student Conceptions of Feedback’ survey instrument with the SRL score and with the GPA score.

9.3 Identification of contribution to knowledge

9.3.1 SRL Conceptual Framework

This thesis presents the development of a conceptual framework for SRL which overtly incorporates the role of reflection which is undertaken by effective self-regulated learners on their performances after an assessment as part of their self-assessment of their abilities before and after engaging with feedback generated during the assessment. The conceptual framework is contextualised to the feedback provided after clinical examinations at Birmingham but could be adapted to other learning contexts, including engagement with feedback after formative as well as summative assessments. This is of wider interest because the conceptual framework makes overt the internal processes and external actions that are required for effective engagement with the feedback. It is believed that students can learn to become more effective self-directed learners (Nicol and Macfarlane Dick 2006, p205; Sandars and Cleary 2011, p875) and so be facilitated to make more effective use of feedback as appropriate. The potential to provide feedback after summative clinical assessments is becoming more feasible as more undergraduate medical courses are using electronic marking systems (Kropmans et al. 2015; Denison et al. 2016), and it is important that managers of healthcare profession courses ensure that all students are equipped to engage with such feedback to improve their knowledge and skills for professional practice.

9.3.2 Engagement With Feedback Questionnaire

The EWF questionnaire was developed over two iterations in order to assess students’ engagement with their feedback by differentiating the processes that comprise effective engagement with feedback, as detailed in the conceptual framework. The results of the Phase 2 administration of the EWF items demonstrate that the majority of students, therefore including all levels of academic performance, reported they engaged with their feedback from the clinical examinations in terms of accessing and cognitively engaging with it, but not as many students reported that they acted on the feedback, even though the majority of students agreed that it is important to act on this feedback. This adds

to the body of literature from undergraduate clinical education and other higher education disciplines which indicate that merely providing feedback is not enough to ensure that it is acted on, even with regard to a course such as this where the feedback relates to knowledge and skills that could be required in future professional practice (Bing-You 1997; Sargent et al. 2009; Harrison et al. 2014). Importantly, what these results highlight is that in this context, whole cohort interventions to improve students' propensity to act in order to remedy identified deficiencies in knowledge and skills, would be appropriate rather than focussed interventions for students at lower performance levels.

9.3.3 Reflection Questionnaire

The results presented in this thesis add to what has already been published about Kember et al.'s (2000) Reflection Questionnaire. The results for internal reliability (Cronbach's alpha) in Phase 1 & 2 and validity in Phase 2 were very similar to administrations of the survey in other healthcare related disciplines. Tricio et al. (2015) also administered the Reflection Questionnaire and compared the results for the Reflection and Critical reflection sub-scales with the results from individual types of assessments. Their results also indicated no association between the two RQ sub-scales and the OSCE scores. Grant et al. (2006) and Lucieer et al. (2016) conducted similar studies which used other scales and similarly failed to demonstrate an association between students' disposition to reflect on learning and their academic performance. The reasons why there does not appear to be an association between skills at reflection on learning and academic performance may be because these studies are all conducted with healthcare students who have been selected because they have these skills, or have been given opportunities to enhance these skills during their course of study.

9.4 Further research on this topic

The next steps in researching this topic could include investigating:

Self-regulated learning - Self-assessment of performance

The ability of self-regulated learners to reflect on and self-assess performance after an assessment is an important aspect which has not been extensively explored with regard to a SRL framework. An interview study could be devised to

examine how well students at different levels of academic performance were able to self-assess their performance in each question in a clinical examination with regards to the specific knowledge and skills that were assessed (rather than focussing on whether they felt they had passed or failed the question) *before* the release of the feedback, and later, how much congruence this self-assessment had with the strengths or weaknesses highlighted in the subsequent examiner feedback comments. The findings from this could then be used to provide support to students in developing SRL ability to self-assess progress towards specific learning goals or externally set learning outcomes, perhaps by including the use of peer formative assessment activities.

What does 'acts on the feedback' mean?

There was no attempt in the survey studies to define what 'acts' means in practice. Definitions given by students in the interview study ranged from asking a friend what they thought the right answer to a question was, looking up facts on Google or in relevant textbooks, to practising, or planning to practise, clinical skills in simulation or with patients. It would be useful to investigate what students at different levels of performance define as 'acting on the feedback' and what their most common approaches to action are in order to explore how students might be encouraged or facilitated to maximise the effectiveness of what they choose to act on and how they achieve these learning goals. The findings from this could then be used to provide support for students to aim for excellence in all competencies, rather than accepting that they demonstrated minimal passing standards in the assessment (Harrison et al., 2014). This should be supported by guidance in making appropriate action plans e.g. using a SMART planning approach (Doran, 1981).

Quality of the individual feedback comments

The quality of the feedback provided has not been explored in any depth for this thesis. The small sample of students in the interview study did not report major difficulties in legibility, their ability to understand the feedback or in knowing what they should do in order to act on the feedback. Obviously the quality of the feedback could have a major impact on motivating or guiding students to act on it if, for example it is not legible or sufficiently specific. A future avenue of research

could be to conduct a structured analysis of a representative sample of feedback comments. There are published papers on how to define the quality of written feedback via metric based frameworks (e.g. White and Sharma 2012; Hughes et al. 2015; Denison et al. 2016; Bartlett et al. 2017). The findings from this could be fed back to examiners and examples of good and poor practice anonymised for use in faculty development.

9.5 Recommendations for this context

Development of a feedback culture at Birmingham

Archer (2010) and Watling (2014) both discuss the notion of a feedback culture whereby systems, structures and processes are developed within an educational context to ensure that the feedback provided is used as feedforward and the cultural assumption is that feedback will routinely be acted on. Archer asserts that 'The artificial dichotomy between summative and formative feedback often distracts from generating potentially powerful feedback from all assessment' (2010, p106). However, detailed feedback is currently provided after every type of summative assessment on the MB ChB programme at Birmingham, as recommended in Norcini et al. (2011).

As noted in Chapter 1, the annual National Student Survey (NSS) had a role in driving improvements to the feedback provided to students on this course, and to students in higher education nationally. The NSS scores for this course in 2016 show a marked improvement compared to 2013 for the items on assessment and feedback (Appendix 8). These scores could be further improved if Archer's advice was implemented. The results of the NSS are a publically available metric and so they are important to all HE institutions from a management perspective (Richardson et al. 2007). However, the potential benefits to future patients which could be derived from healthcare professionals who graduate with a grounding in the skills to engage effectively with feedback present a more powerful argument for the implementation of a feedback culture.

There have been actions to improve the support and guidance for students on the Birmingham Medics course with regard to feedback, but currently there is no overarching vision to inculcate a course wide 'feedback culture'. For example:

- Y1 and 2 students are advised to use a pro forma which guides their reflection on assessments and the resulting feedback. There is a similar pro forma for Y3 students for the year-long clinical module activities.
- Y4 students who failed or were a borderline pass in their previous end of year clinical examination are encouraged to use a mini-CEX pro forma (Norcini, 2005) to gain feedback on their clinical skills from their peers or their clinical supervisors.
- A 'Bank of Assessed Work' has been developed and hosted on the University's virtual learning environment. This includes the examiner's information and mark sheets for exemplar clinical examination questions along with filmed clips of volunteer students role playing performances at different standards in order to improve students' understanding of the marking processes and the expected standards.
- The feedback comments from the Y3, 4 and 5 end of year clinical examinations in spring 2017 were rebadged as 'feed forward' to highlight that action is expected. I amended the headers on the mark sheet pro forma from 'What did you do well?' and 'What could you improve?' to 'Feedback on what you did well' and 'Feedforward on what you should act on to improve'. In the Clinical Examiners' training sessions in March 2017, I included guidance on this, and asked examiners to frame their feedback comments as what was done well in the OSCE and what should be improved for future clinical practice.
- Two sessions of the same faculty development workshop on 'How to give effective feedback' were run in March 2015 and were oversubscribed. The evaluations from the clinical tutors who participated indicated that the information in the workshop was well received.

However, in support of the concept of an assessment culture, Archer believes that 'Feedback must be conceptualised as a supported sequential process rather than a series of unrelated events' (2010, p106).

Suggestions on how a course-long feedback culture could be developed

These recommendations may be applicable to other learning contexts.

- Develop students' SRL skills to improve their engagement with feedback, especially with regard to aspects of accurate self-assessment and reflection on their learning. As recommended by Bing-You and Trowbridge (2009) training in how to recognize, receive, and respond to feedback at a metacognitive level could be included in the curriculum via interactive sessions in each year of the course to highlight the importance of reflecting and acting on feedback provided as part of the students' normal learning routines; not only after assessments, but also to empower them to ask for feedback from peers and tutors so they are at the centre of the feedback process (Rudland et al., 2013) and are co-producers of learning (Boud and Molloy, 2013). This would potentially mitigate against some of the structural issues of learning in a clinical workplace, with frequent rotations between sites, medical specialities and tutors.

- Introduce the structured use of peer formative assessment in each year of the course. It has been found that students who provide qualitative, constructive feedback to their peers about their learning activities are able to reflect on and improve their own learning because they have a better understanding of the expected standards and where their own strengths and deficiencies lie (Perera et al. 2010).

- Provide an e-portfolio so students can map and reflect on their progress over the duration of the whole course (Driessen et al. 2005; Driessen et al. 2007; Buckley et al. 2009; Beckers et al. 2016). Embedded in the portfolio approach to encouraging reflection on progress is the need to develop students' skills in making effective action plans to improve knowledge or skills as required. Students could also be required to provide evidence in their portfolio about how they have reflected and acted on their feedback.

- Improve Faculty skills in the provision of face to face and written feedback. Hold regular faculty development workshops to ensure that the large and changing pool of tutors and examiners have at least baseline knowledge of

the principles of, and practice in giving, effective verbal and written feedback in order to optimise the use of feedback as a learning tool (Bynum 2015; Molloy 2010). It could be argued that changes embedded at the micro-level of the individual tutor and their educational interactions with students have the potential to make the greatest difference to the success of inculcating a feedback culture. However, the effects of the constraints of the pressured healthcare environment which clinical tutors work in should not be underestimated.

For an assessment culture to exist, these suggested activities and others would need to be cross referenced, understood and supported by all the actors; students, tutors, course leaders and administrators. The activities would need to be evaluated and updated as appropriate to maintain their relevance. It has been argued that an effective way to ensure a feedback culture is to implement a process of 'programmatic assessment' (Schuwirth and van der Vleuten, 2011; Eva et al. 2016). This entails moving away from the current dominance of summative assessment methods which attempt to maximise reliability and instead recognise the importance of using assessment methods in which validity and authenticity have primacy. Examples are work-based methods such as Mini-CEX (Norcini, 2005) which provide many low-level formative assessments with feedback and the expectation that this feedback will be acted on to improve performance at the next assessment, with the qualifying award being mainly based on this evidence.

9.6 Author's reflection on this thesis

In Chapter 4, I claimed to be a reflective practitioner (Schön, 1991) and in support of this, include here a brief reflection on the learning journey this thesis has afforded me. In Chapter 4 I also claimed to endorse theory that informs effective practice. This point is key to my rationale for undertaking this doctorate. I wanted to study, in depth, an educational phenomenon that was relevant to my professional context in order to improve some part of the educational experience of the learners in my context and my professional practice and to potentially generate results that could have wider applicability in undergraduate medical education.

Although I am able to describe the rationale for the three empirical studies which are presented here and account for the methodological choices made, my research journey was much more circuitous than the linear one described in text books such as Thomas (2013). Once I had chosen my topic, I knew the first thing I needed to do was to talk to some students and ask them what (if anything) they did with the feedback comments to explore if the initial outline conceptual framework would be of use. It was a pleasure to interview the students, and I felt that their frank views gave me a window into their world which was a really helpful starting point. Because the narratives they gave me were so interesting and powerful in terms of credibility and internal cohesion, I needed to remind myself that the data from these interviews could be viewed as anecdotal in terms of the strength of evidence that could be applied to the potential activities of the wider student population.

I had originally envisaged undertaking an evaluation survey to find out more about what students thought about and did with the feedback. I was advised that a survey approach that sought to measure latent constructs, beliefs, or propensity to engage with feedback would be a more fruitful approach because of the opportunities for establishing the reliability and validity of the data via statistical analysis (including factor analysis) and the potential to generate results that could be generalised to other healthcare education settings. Although I agreed with this, I was not fully divorced from the quest to understand more about what the learners in my context did with this specific, resource intensive feedback, rather than what they thought about or did with feedback in general. Because of this, the Engagement With Feedback (EWF) questionnaire items were not completely focussed on traits and I took a hybrid approach that that may not have served either function effectively. I initially assumed that there would be an appropriate published survey on engagement with feedback in academic situations which I could adapt for this situation; I did not expect to develop one de novo. I did not allocate sufficient time to do this properly by including external verification steps from, for example, a Delphi group approach (Bloor et al. 2015). After I conducted these studies, Brown et al. (2016) published their report on the administration of their ‘Student conceptions of feedback’ scale which would have been an appropriate scale to adapt for my studies.

I am not confident in statistical analytical methods and I followed guidance in Pallant (2013) and other text books, sometimes without having a real understanding of the mathematical processes involved in, for example, factor analysis, despite carefully reading books like Pett (2003) which explain the underlying principles of the data rotation methods. I undertook many different factor analyses, using different approaches because I wanted to be certain that my results were correct and I had not selected the wrong method or followed the processes in IBM SPSS incorrectly. As suggested by Pallant (2013, p205), these different approaches to factor analysis did show similar results, which was reassuring.

The research questions I asked led me to use both qualitative and quantitative methods to garner data. This increased the value of this research journey to me as I now have experience of both research approaches and their strengths and weaknesses (and mine too!). The experience of developing the EWF sub-scales over two iterations has enabled me to develop skills which enhance my understanding when reading reports about the development and administration of surveys and the credibility of results from these. All the skills that I have learned during this research apprenticeship will be invaluable in future endeavours.

9.7 Summary

This thesis discusses the development of a conceptual framework for self-regulated engagement with feedback after a clinical examination. This chapter reminds the reader of the context of the three empirical studies and the research questions asked in this thesis. The findings from the three empirical studies are summarised. The areas of advances in knowledge are highlighted and suggestions made for specific areas of further study to build on these findings. Recommendations are presented for improvements to this course to benefit learners by improving their skills in acting on feedback, which have resonance for other professional courses. Finally, I briefly reflect on this personal learning journey.

REFERENCES

Ablard, K., E. and Lipschultz, R. E (1998) 'Self-Regulated Learning in High-Achieving Students: Relations to Advanced Reasoning, Achievement Goals, and Gender', *Journal of Educational Psychology*, 90(1), pp. 94-101

Agiusa, N.M. and Wilkinson, A. (2014) 'Students' and teachers' views of written feedback at undergraduate level: A literature review', *Nurse Education Today,* 34 (4), pp. 552–559

Ajjawi, R (2012) 'Gong beyond 'received and understood' as a way of conceptualising feedback', *Medical Education*, 46, pp 1017-1019

Ajjawi, R. and Boud, D (2017) 'Researching feedback dialogue: an interactional analysis approach' *Assessment and Evaluation in Higher Education*, 42(2), pp 252-265

Archer, J. (2010) 'State of the science in health professional education: effective feedback', *Medical Education,* 44, pp. 101–108. DOI: 10.1111/j.1365-2923.2009.03546.x

Aronson, L., Niehaus, B., Hill-Sakurai, L., Lai, C. and O'Sullivan, P. (2012). 'A Comparison of Two Methods of Teaching Reflective Ability in Year 3 Medical Students', *Medical Education,* 46, pp. 807-814

Ashby, S., Snodgrass, S., Rivett, D., Russell, T. (2016) 'Factors shaping e-feedback utilization following electronic objective structured clinical examinations', *Nursing & Health Sciences*, 18, pp. 362-369

Ashford S.J. (1986) 'Feedback-seeking in individual adaptation: a resource perspective', *Academic Management Journal,* 29 pp. 465–87

Aukes, L.C., Geertsma, J., Cohen-Schotanus, J., Zwierstra, R.P., Slaets, J.P.J. (2007) 'The development of a scale to measure personal reflection in medical practice and education' *Medical Teacher*, 29(2-3) pp. 177-182, DOI: 10.1080/01421590701299272

Bailey, R. and Garner, M. (2010) 'Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices', *Teaching in Higher Education*, 15(2), pp. 187-198 http://dx.doi.org/10.1080/13562511003620019

Baker, M (2016) *Nature* 25-5-2016 accessed 2-2-2017http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

Bandura, A. (1977) *Social learning theory*. Englewood Cliffs ; London : Prentice-Hall

Bandura, A. & Cervone, D. (1983). 'Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems', *Journal of Personality and Social Psychology*, 45, pp. 1017–1028

Barr, R.B. and Tagg, J. (1995) 'From teaching to learning – a new paradigm for undergraduate education', Change: The Magazine of Higher Learning, 27(6), pp. 12-26,

Bartlett, M., Crossley, J., McKinley, R. (2017) 'Improving the quality of written feedback using written feedback', *Education for Primary Care*, 28(1), pp. 16-22. DOI: 10.1080/14739879.2016.1217171

Baumgardner, A. H., Kaufman, C.M and Levy, P. E (1989) 'Regulating Affect Interpersonally: When Low Esteem Leads to Greater Enhancement', *Journal of Personality & Social Psychology,* 56(6) pp. 907-921

Beckers, J., Dolmans, D.H.J.M., & van Merriënboer, J.J.G. (2016). 'e-Portfolios enhancing students' self-directed learning: A systematic review of influencing factors', *Australasian Journal of Educational Technology*, 32(2), pp. 32-46.

BERA Ethical Guidelines for Educational Research (2011) London: British Educational Research Association
https://www.bera.ac.uk/wp-content/uploads/2014/02/BERA-Ethical-Guidelines-2011.pdf (Accessed 19-2-2017)

Beaumont, C., O'Doherty, M., Shannon, L. (2011) 'Reconceptualising assessment feedback: a key to improving student learning?', *Studies in Higher Education*, 36(6), pp. 671-687

Biggs, J. (1996) 'Enhancing teaching through constructive alignment', *Higher Education* 32, pp. 347-364

Bing-You, R.G., Paterson, J. and Levine, M.A. (1997) 'Feedback falling on deaf ears: residents' receptivity to feedback tempered by sender credibility', *Medical Teacher*, 19(1) pp. 40-44.

Bing-You RG, Trowbridge RL. (2009) 'Why Medical Educators May Be Failing at Feedback', *JAMA,* 302(12), pp. 1330-1331. doi:10.1001/jama.2009.1393

Black, N.M.I. and Harden, R.M. (1986) 'Providing feedback to students on clinical skills by using the Objective Structured Clinical Examination', *Medical Education*, 20, pp. 48-52

Black, P. and Wiliam, D (1998) *Inside the black box*: *Raising standards through classroom assessment,* School of Education, King's College, London, United Kingdom.

Blair, A., Wyburn-Powell, A., Goodwin, M. and Shields, S. (2014) 'Can dialogue help to improve feedback on examinations?', *Studies in Higher Education*, 39(6), pp.1039–1054, http://dx.doi.org/10.1080/03075079.2013.777404

Bloor, M., Sampson, H., Baker, S., Dahlgren, K. (2015) 'Useful but no Oracle: reflections on the use of a Delphi Group in a multi-methods policy research study', *Qualitative Research,* 15, pp. 57–70. DOI: 10.1177/1468794113504103

Bok H.G.J., Teunissen, P.W., Favier, R.P., Rietbroek, N.J., Theyse, L.F.H., Haarhuis, J.C.M., van Beukelen, P., van der Vleuten, C.P.M., Jaarsma, D.A.D.C. (2013) 'Programmatic assessment of competency-based workplace learning: when theory meets practice', *BMC Med Educ;*13 *(1),* 123 pp.1-9

Boud, D. (1995) *Enhancing Learning through Self-Assessment,* London: Kogan Page

Boud, D. (2015) 'Feedback: ensuring that it leads to enhanced learning', *The Clinical Teacher,* 12, pp. 3-7.

Boud, D. and Falchikov, N. (2007) 'Introduction: assessment for the longer term' in Boud, D. and Falchikov, N. (Eds) *Rethinking Assessment in Higher Education: Learning for the Longer Term*. London: Routledge pp. 3-13

Boud, D., Lawson, R. and Thompson, D.G. (2013) 'Does student engagement in self-assessment calibrate their judgement over time?', *Assessment and Evaluation in Higher Education*, 38(8), pp. 941-956

Boud, D. Lawson, R. and Thompson, D.G. (2015) 'The calibration of student judgement through self-assessment: disruptive effects of assessment patterns', *Higher Education Research & Development*, 34(1), pp. 45-59.

Boud, D and Molloy, E. (2013) 'Rethinking models of feedback for learning: the challenge of design', *Assessment and Evaluation in Higher Education*, 38(6) pp. 698-712

Boud, D and Molloy, E. (2013) 'What is the problem with feedback?' in: Boud D, and Molloy E. (Eds) *Feedback in Higher and Professional Education*. Abingdon:Routledge pp.1-10

Boud, D. and Walker, D. (1998). 'Promoting reflection in professional courses: The challenge of context', *Studies in Higher Education*, 23(2), pp.191-206.

Bounds, R., Bush, C., Aghera, A., Rodriguez, N., Stansfield, R.B., Santen, S.A. (2013) 'Emergency Medicine Residents' Self assessments Play a Critical Role When Receiving Feedback', *Academic Emergency Medicine,* 20(10) pp. 1055-1061

Boursicot, K.A.M., Roberts, T.E. and Burdick, W.P. (2010) 'Structured assessments of clinical competence' in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education, pp.246-258

Bowen, L., Marshall, M., Murdoch-Eaton, D. (2017) 'Medical student perceptions of feedback and feedback behaviours within the context of the 'educational alliance'', *Academic Medicine,* March 2017 ahead of print DOI: 10.1097/ACM.0000000000001632

Braun, V. and Clarke, V. (2008) 'Using thematic analysis in psychology', *Qualitative Research in Psychology*, 3(2), pp.77-101

Braun, V. and Clarke, V. (2013) *Successful Qualitative Research* London Sage

Bristol Online Survey (BOS) https://www.onlinesurveys.ac.uk Accessed date 28-10-2016

Broadfoot, P. M., Daugherty, R., Gardner, J., Harlen, W., James, M., & Stobart, G. (2002) *Assessment for learning: 10 principles*. Cambridge, UK: University of Cambridge School of Education

Brown, G.T.L, Peterson, E.R., Yao, E.S. (2016) Student conceptions of feedback: Impact on self-regulation, self-efficacy, and academic achievement, *British Journal of Educational Psychology*, 86, pp. 606-629 DOI:10.1111/bjep.12126

Brydges, R. and Butler, D. (2012) 'A reflective analysis of medical education research on self-regulation in learning and practice', *Medical Education,* 46(1) pp.71-79.

Bryman, A. (2006) 'Paradigm Peace and the Implications for Quality ', *Int. J. Social Research Methodology* 9(2), pp. 111–126

Buckley, S et al (2009) 'The educational effects of portfolios on undergraduate student learning: a Best Evidence Medical Education (BEME) systematic review. BEME Guide No. 11', *Medical Teacher*, 31(4), pp282-98. doi: 10.1080/01421590902889897.

Bunniss, R and Kelly D.R. (2010) 'Research paradigms in medical education research', *Medical Education,* 44, pp. 358-366

Burgess, H., Sieminski, S., and Arthur, L. (2006) *Achieving Your Doctorate in Education* London:SAGE Publications Ltd

Butler, D. L. & Winne, P. H. (1995) 'Feedback and self-regulated learning: a theoretical synthesis', *Review of Educational Research*, 65(3), pp. 245–281.

Bynum, W.E. (2015) 'Filling the feedback gap: The unrecognised roles of shame and guilt in the feedback cycle', *Medical Education*, 49, pp. 644–647.

Cantillon, P., and Sargeant, J (2008), 'Giving feedback in clinical settings', *BMJ,* 337, doi: https://doi.org/10.1136/bmj.a1961

Carifio, J. and Perla, R. (2008) 'Resolving the 50-year debate around using and misusing Likert scales', *Medical Education,* 42 pp.1150–1152.

Carless, D (2015) *Excellence in University Assessment: Learning from Award-winning Practice, Oxon:* Routledge

Carruthers, C., McCarron, B., Bolan, P., Devine, A. McMahon-Beattie, U. and Burns, A. (2015) ''I like the sound of that' – an evaluation of providing audio feedback via the virtual learning environment for summative assessment', *Assessment & Evaluation in Higher Education*, 40(3), pp.352–370, http://dx.doi.org/10.1080/02602938.2014.917145

Catell, R.B. (1966) 'The scree test for number of factors', *Multivarate Behavioural Research,* 1 pp.245-76

Clark, I. (2012) 'Formative Assessment: Assessment is for Self-regulated Learning', *Educational Psychology Review,* 24, pp. 205-249

Clark and Watson (1995) 'Constructing Validity: Basic Issues in Objective Scale Development', *Psychological Assessment,* 7(3), pp.309-319.

Cleary, T.J., Callan, G.L., Zimmerman, B.J. (2012) 'Assessing Self-Regulation as a Cyclical, Context-Specific Phenomenon: Overview and Analysis of SRL Microanalytic Protocols', *Education Research International, 2012,* pp. 1-19 http://dx.doi.org/10.1155/2012/428639

Colthart, I., Bagnall, G., Evans, A., Allbutt, H., Haig, A., Illing, J. and McKinstry, B. (2008) 'The effectiveness of self-assessment on the identification of learner needs, learner activity, and impact on clinical practice: BEME Guide no. 10', *Medical Teacher,* 30, pp. 124–145

Costello, A.B. and Osborne, J.W (2005) 'Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis', *Practical Assessment, Research and Evaluation*, 10(7), pp. 1-9

Creswell, J.W., (2014) *Research Design*, Thousand Oaks Cal: SAGE Publications Inc.

Crommelinck, M. and Anseel, F. (2013), 'Understanding and encouraging feedback-seeking behaviour: a literature review', *Medical Education*, 47, pp. 232–241. doi:10.1111/medu.12075

Cronbach L.J. (1951) 'Coefficient alpha and the internal structure of tests', *Psychometrika,*16, pp. 297–334

Davis, D.A., Mazmanian, P.E., Fordis, M., Van Harrison, R., Thorpe, K.E., Perrier, L. (2006) 'Accuracy of physician self-assessment compared with observed measures of competence: A systematic review' *JAMA* 296(9), pp. 1094-1102

Denison, A., Bate, E., Thompson, J. (2016) 'Tablet versus paper marking in assessment: feedback matters', *Perspectives in Medical Education*, 5(2) pp.108-13. doi: 10.1007/s40037-016-0262-8.

Dennick, R. (2012) 'Twelve tips for incorporating educational theory into teaching practices'**,** *Medical Teacher*, 34(8), pp.618-624

deVaus, D. (2014) *Surveys in Social Research 6th edn*. Oxon: Routledge

DeVellis, R.F. (2017) *Scale Development: Theory and Applications 4th Edition*, Los Angeles: SAGE

DiStefano, C., Zhu, M., Mindrila, D. (2009) 'Understanding and Using Factor Scores:Considerations for the Applied Researcher', *Practical Assessment, Research & Evaluation*, 14(20), pp. 1-11

Doran, G. T. (1981). 'There's a S.M.A.R.T. way to write management's goals and objectives', *Management Review,* AMA FORUM, 70(11), pp. 35–36

Driessen, E., Van Tartwijk, J., Overeem, K., Vermunt, J.D., van der Vleuten, C (2005) 'Conditions for successful reflective use of portfolios in undergraduate medical education', *Medical Education*, 39(12), pp.1230-5.

Driessen, E., Van Tartwijk, J., van der Vleuten, C., Wass V. (2007), 'Portfolios in medical education: why do they meet with mixed success? A systematic review', *Medical Education*, 41, pp1224–1233

Dweck, C.S. (1999) *Self-theories: their role in motivation, personality, and development* Philadelphia, PA; Hove: Psychology Press

Dweck, C.S., Leggett E.L. (1988) 'A social-cognitive approach to motivation and personality', *Psychological Review*, 95 (2), pp. 256–273

Ende, J. (1983) 'Feedback in Clinical Medical Education', *JAMA*, 250(6), pp. 777-781

Epstein, R.M., Siegel, D.J., Silberman, J. (2008) 'Self-monitoring in clinical practice: A challenge for medical educators', *Journal of Continuing Education in the Health Professions*, 28(1), pp.5-13

Ertmer P.A. and Newby T.J. (1993) 'Behaviourism, Cognitivism, Constructivism: Comparing Critical Features From an Instructional Design Perspective' *Performance Improvement Quarterly*, 6(4), 1993, pp. 50–72. doi: 10.1111/j.1937- 8327.1993.tb00605.x

Ertmer, P.A. and Newby, T.J. (1996) 'The expert learner: Strategic, self-regulated and reflective', *Instructional Science* 24, pp.1-24

Eva, K.W., Munoz, J., Hanson, M.D., Walsh, A., Wakefield, J. (2010), 'Which factors, personal or external, most influence students' generation of learning goals?', *Academic Medicine,* 85(10 Suppl), S pp.102-5.

Eva, K.W., Armson, H., Holmboe, E., Lockyer, J., Loney, E., Mann, K., Sargeant, J. (2012) 'Factors influencing responsiveness to feedback: on the interplay between fear, confidence and reasoning processes', *Advances in Health Science Education*, 17, pp. 15-26

Eva, K.W, Bordage, G., Campbell, C., Galbraith, R., Ginsburg, S., Holmboe, E., Regehr, G. (2016) 'Towards a program of assessment for health professionals: from training into practice', *Advances in Health Sciences Education*; 21(4), pp.897-913 DOI 10.1007/s10459-015-9653-6

Evans, C. (2013) 'Making sense of assessment feedback in higher education', *Review of Educational Research,* 83(1) pp. 70-120

Falchikov, N. (2007) 'The place of peers in learning and assessment' in Boud, D. and Falchikov, N. (Eds) *Rethinking assessment in higher education: learning for the longer term*. London: Routledge pp. 128-143

Fedor, S.M., Rensvold, D.B., Adams, R.B. (1992) 'An investigation of factors expected to affect feedback seeking: a longitudinal field study', *Personnel Psychology* 45, pp. 779–780

Fishbach, A., Eyal, T., Finkelstein, S.R., (2010) 'How positive and negative feedback motivate goal pursuit,' *Social and Personality Psychology Compass*, 4(8), pp. 517-530

General Medical Council (2015) *Promoting excellence: standards for medical education and training*
http://www.gmc-uk.org/Promoting_excellence_standards_for_medical_education_and_training_0715.pdf_61939165.pdf (accessed 31-1-2017)

Gibbs, G. and Simpson, C (2004) 'Conditions Under Which Assessment Supports Students' Learning', *Learning and Teaching in Higher Education*, 1, pp. 3-31

Gierl, M.J., Latifi, S., Lai, H., Philippe Boulais, A., De Champlain, A (2014) 'Automated essay scoring and the future of educational assessment in medical education', *Medical Education*: 48 pp. 950–962

Glaser BG, Strauss AL. (1967) *The Discovery of Grounded Theory: Strategies for Qualitative Research* Chicago, IL: Aldine

Gordon MJ. (1991) 'A Review of the Validity and Accuracy of Self-assessments in Health Professions Training', *Academic Medicine*, 66, pp. 762–769

Grant, A., Kinnersley, P., Metcalf, E., Pill, R. Houston, H. (2006) 'Students' views of reflective learning techniques: an efficacy study at a UK Medical School', *Medical Education* 40, pp. 379-388

Grant, H. and Dweck, C.S. (2003) 'Clarifying achievement goals and their impact', *Journal of Personality and Social Psychology*, 85(3), pp. 541-553

Hammersley, M (2001) 'On 'systematic' reviews of research literatures: a 'narrative' response to Evans and Benefield', *British Educational Research Journal*, 27(5), pp. 543-554

Handley, K., Price, M. and Millar, J. (2011) 'Beyond 'doing time': investigating the concept of student engagement with feedback', *Oxford Review of Education*, 37(4), pp. 543-560

Harden R., Gleeson F.A. (1979) 'Assessment of clinical competence using an objective structured clinical examination (OSCE)', *Medical Education,* 13, pp. 41-45.

Harden, R.M., Grant, J. and Hart, I.R. (1999). 'BEME Guide No. 1:Best Evidence Medical Education'**,** *Medical Teacher* 21(6) pp.553-562

Harrison, C. J., Könings, K. D., Molyneux, A., Schuwirth, L. W., Wass, V., and van der Vleuten C. P. (2013) 'Web-based feedback after summative assessment: how do students engage?,' *Medical Education,* 47(7), pp. 734-744. DOI: 10.1111/medu.12209

Harrison, C. J., Könings, K. D., Schuwirth, L., Wass, V and van der Vleuten C.P. (2014) 'Barriers to the uptake and use of feedback in the context of summative assessment', *Advances in Health Sciences Education* 20(1), pp. 229-245.

Harrison, C. J., Molyneux, A.J., Blackwell, S. and Wass, V.J. (2015) 'How we give personalised audio feedback after summative OSCEs', *Medical Teacher*, 37, pp. 323–326 DOI: 10.3109/0142159X.2014.932901

Hartley, J. and Betts, L. (2010) 'Four layouts and a finding: the effects of changes in the order of the verbal labels and numerical values on Likert-type scales', *International Journal of Social Research Methodology,* 13(1), pp. 17-27.

Hattie, J., and H. Timperley (2007), 'The power of feedback', *Review of Educational Research* 77, pp. 81–112.

Health Education England (2016) 'Improving safety through education and training. Report by the Commission on Education and Training for Patient Safety' Available at https://hee.nhs.uk/sites/default/files/documents/FULL%20report%20medium%20res%20for%20web.pdf (accessed 28-5-2017

Heeneman, S., Pool, A.O., Schuwirth, L.W.T., van der Vleuten, C.P.M., Driessen, E.W. (2015) 'The impact of programmatic assessment on student learning: theory versus practice', *Medical Education,* 49(5), pp.487–98

Henderson, P., Ferguson-Smith, A.C., and Johnson, M.H. (2005) 'Developing essential professional skills: a framework for teaching and learning about feedback', *BMC Medical Education*, 5(11), pp. 1-6 DOI: 10.1186/1472-6920-5-11

Hepplestone, S., Holden, G., Irwin, B., Parkin, H.J., Thorpe, L. (2011) 'Using technology to encourage student engagement with feedback: a literature review', *Research in Learning Technology*, 19(2), pp. 117-127

Hepplestone, S., Chikwa, G. (2014) 'Understanding how students process and use feedback to support their learning', *Practitioner Research in Higher Education Journal*, 8(1), pp. 41-53.

Hewson, M.G. & Little, M.L. (1998) 'Giving feedback in medical education', *J GEN INTERN MED,* 13(2) pp.111-116

Higgins, R., Hartley, P., and Skelton, A. (2002) 'The Conscientious Consumer: reconsidering the role of assessment feedback in student learning', *Studies in Higher Education,* 27(1), pp. 53-64 http://dx.doi.org/10.1080/03075070120099368

Hodges, B. (2003) 'OSCE! Variations on a Theme', *Medical Education*; 37, pp. 1134–1140

Holmes, N. (2015) 'Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module', *Assessment & Evaluation in Higher Education*, 40(1), pp. 1–14, http://dx.doi.org/10.1080/02602938.2014.881978

Horn, J.L., (1965) 'A rationale and test for the number of factors in factor analysis', *Psychometrika*, 30, pp.179-85

Hughes, G., Smith, H., Creese, B. (2015) 'Not seeing the wood for the trees: developing a feedback analysis tool to explore feed forward in modularised Programmes', *Assessment & Evaluation in Higher Education*, 40(8) pp.1079-1094, DOI:10.1080/02602938.2014.969193

Humphris, G M and Kaney, S. (2001) 'Examiner fatigue in communication skills objective structured clinical examinations', *Medical Education*, 35, pp. 444–449 10.1046/j.1365-2923.2001.00893.x

IBM Corp. (2013) IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Irwin, B., Hepplestone, S., Holden, G., Parkin, H.J., Thorpe, L. (2013) 'Engaging students with feedback through adaptive release', *Innovations in Education and Teaching Internationa*l, 50(1), pp. 51-61

Jamieson, S. (2004), 'Likert scales: how to (ab)use them'. *Medical Education,* 38, pp. 1217–1218.

Jolly, B. and Boud, D (2013) 'Written feedback: What is is good for and how can we do it well?' in: Boud D, and Molloy E. (Eds) *Feedback in Higher and Professional Education*. Abingdon:Routledge pp. 104-124

Kahn, J.H. (2006) 'Factor Analysis in Counselling Psychology Research, Training, and Practice: Principles, Advances, and Applications' *The Counselling Psychologist*, 34(5), pp. 684-718

Kalk, K., Luik, P., Taimalu, M., and Taht, K. (2014) 'Validity and reliability of two instruments to measure reflection: a confirmatory study', *TRAMES* 18(68/63), 2, pp. 121-134.

Kane, M. T. (2013) 'Validating the Interpretations and Uses of Test Scores', *Journal of Educational Measurement*, 50(1), pp. 1-73. doi:10.1111/jedm.12000

Kaufman, D.M. and Mann, K. V. (2010) 'Teaching and Learning in Medical Education: how theory can inform practice', in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education, pp.16-36

Kember, D., Leung, D., Jones, A., Loke, A.Y., McKay, J., Sinclair, K., Tse, H., Webb, C., Kam, F., Wong, Y., Wong, M., Yeung, E. (2000) 'Development of a Questionnaire to Measure the Level of Reflective Thinking', *Assessment & Evaluation in Higher Education*, 25(4), pp.381-395.

Ker, J. and Bradley, P. (2010) 'Simulation in Medical Education', in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education, pp. 164-180

Khan, K.Z., Ramachandran, S., Gaunt, K., and Pushkar , P. (2013a) 'The Objective Structured Clinical Examination (OSCE): AMEE Guide No. 81. Part I: An historical and theoretical perspective' *Medical Teacher*, 35(1), pp.e1437-1446

Khan, K.Z., Ramachandran, S., Gaunt, K., and Pushkar , P. (2013b) 'Organisation and Administration', *Medical Teacher*, 35(1), pp.e1447–e1463

Kirschner, P., Sweller, J. & Clark, R. (2006) 'Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching'. *Educational Psychologist,* 41**,** pp. 75-86.

Kluger, A.N. and DeNisi, A. (1996) 'The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory', *Psychological Bulletin*. 119. pp. 254-284

Kluger, A.N. and Van Dijk, D. (2010) 'Feedback, the various tasks of the doctor, and the feedforward alternative', *Medical Education*, 44(12) pp. 1166-1174

Kolb, D.A. (1984): *Experiential learning: experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall

Kropmans, T.J.B., Griffin, L., Cunningham, D., Walsh, D., Setyonugroho, W., et al. (2015)' Back to the Future: Electronic Marking of Objective Structured Clinical Examinations and Admission Interviews Using an Online Management Information System in Schools of Health Sciences', *Journal of Health and Medical Informatics*, 6(1), pp. 1-6

Kuper, A., Reeves, S., Levinson, W. (2008a) 'An introduction to reading and appraising qualitative research', *British Medical Journal*, 337:a288 DOI: https//doi.org/10.1136/bmj.a288

Kuper, A., Lingard, L., Levinson, W. (2008b) 'Critically appraising qualitative research' *British Medical Journal,* 337:a1035 DOI: https//doi.org/10.1136/bmj.a1035

Kvale, S. (1996) *Interviews: An Introduction to Qualitative Research Interviewing*, London: Sage.

Ladyshewsky, R.K., (2013) 'The role of peers in feedback processes' in Boud, D and Molloy, E (Eds) *Feedback in Higher and Professional Education.* Abingdon: Routledge pp174-189

Larsen, D. P., Butler, A. C., Roediger, H. L. (2008) 'Test enhanced learning in medical education', *Medical Education*, 42(10), pp. 959-966

Lave, J. and. Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*. New York: Cambridge University Press

Lefroy, J., Watling, C., Teunissen, P.W., and Brand, P. (2015) 'Guidelines: the do's, don'ts and don't knows of feedback for clinical education', *Perspectives in Medical Education*, 4, pp. 284-299

Leggett, H., Sandars, J and Burns, P. (2012) 'Helping students to improve their academic performance: A pilot study of a workbook with self-monitoring exercises', *Medical Teacher,* 34(9) pp. 751-753

Lethbridge, K., Andrusyszyn, M-A., Iwasiw, C., Laschinger, H.K.S. and Fernando, R. (2013) 'Assessing the psychometric properties of Kember and Leung's reflection questionnaire', *Assessment and Evaluation in Higher Education,* 38(3), pp.303-325.

Linderbaum, B.A., and Levy, P.E, (2010) 'The Development and Validation of the Feedback Orientation Scale (FOS)', *Journal of Management,* 36(6) pp1372-1405 DOI: 10.1177/0149206310373145

Lingard, L. and Kennedy, T.J. (2010) 'Qualitative research methods in medical education' in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education pp. 246-258

Livingston, S.A. and Zieky, M.J. (1982) *Passing Scores: A Manual for Setting Standards of Performance on Educational and Occupational Tests.* Princeton, NJ: Educational Testing Services

Lucieer S.M., Jonker, L. Visscher, C., Rikers, R.M.J.P., Themmen, A.P.N. (2016) 'Self-regulated learning and academic performance in medical education'. *Medical Teacher,* 38(6), pp. 585-593.

Mann, K.V. (2011) 'Theoretical perspectives in medical education: past experience and future possibilities', *Medical Education,* 45, pp. 60-68

Mann, K., Gordon, J. & MacLeod, A. (2009) 'Reflection and reflective practice in health professions education: a systematic review', *Advances in Health Science Education*, 14, pp. 595-621. doi:10.1007/s10459-007-9090-2

McConnell, M.M., Regehr, G., Wood, T.J., and Eva K.W. (2012), 'Self-monitoring and its relationship to medical knowledge', *Advances in Health Sciences Education*. 17(3), pp. 311-23.

Messick, S. (1994). 'The interplay of evidence and consequences in the validation of performance assessments', *Educational Researcher*, 23(2), pp. 13-23

Mezirow, J. (1991) *Transformative Dimensions of Adult Learning*. San Francisco, Jossey-Bass

Miller GE. (1990) 'The assessment of clinical skills/competence/performance', *Academic Medicine,* 65, pp. S63–S67

Mitchell, R. (1994) 'The Development of the Cognitive Behavior Survey to Assess Medical Student Learning', *Teaching and Learning in Medicine*, 6, pp. 161-167

Molloy, E.K. (2010) 'The feedforward mechanism: A way forward in clinical learning?' *Medical Education*, 44, pp.1157–1158.

Molloy, E., Borrell-Carrio, F. and Epstein, R.(2013) 'The impact of emotions in feedback' in: Boud D, and Molloy E. (Eds) *Feedback in Higher and Professional Education*. Abingdon:Routledge pp. 50-71

Molloy, E. and Boud, D. (2013) 'Changing conceptions of feedback', in: Boud D, and Molloy E. (Eds) *Feedback in Higher and Professional Education*. Abingdon:Routledge pp. 11–33

Morris, C. and Blaney, D. (2010) 'Work-based learning' in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education, pp. 69-98

Mossey, P.A., Newton, J.P., and Stirrups, D.R. (2001) 'Scope of the OSCE in the assessment of clinical skills in dentistry', *British Dental Journal* 190(6), pp.323-326 doi:10.1038/sj.bdj.4800961

Murdan, S; (2002) 'Feedback from examinations: Returning marked examination scripts to students', *Pharmacy Education*, 2(4) pp. 167-170

Murdoch-Eaton, D and Sargeant, J. (2012) 'Maturational differences in undergraduate medical students' perceptions about feedback' *Medical Education*, 46, p711-721

Nestel, D., Kneebone, R., Nolan, C., Akhtar, K., Darzi, A. (2011) 'Formative Assessment of Procedural Skills: Students' Responses to the Objective Structured Clinical Examination and the Integrated Performance Procedural Instrument', *Assessment and Evaluation in Higher Education* 36, pp. 171-183

Newble and Jaeger (1983) 'The effect of assessments and examinations on the learning of medical students', *Medical Education,* 17, pp. 165-171

Nguyen, Q.D., Fernandez, N., Karsenti, T.; Charlin, B. (2014) 'What is reflection? A conceptual analysis of major definitions and a proposal of a five-component model', *Medical Education*, 48(12), pp.1176-1189.

Nicol, D. (2010), 'From monologue to dialogue: improving written feedback processes in mass higher education', *Assessment & Evaluation in Higher Education* 35(5), pp. 501-517.

Nicol, D. J. and Macfarlane-Dick, D. (2006) 'Formative assessment and self-regulated learning: a model and seven principles of good feedback practice', *Studies in Higher Education,* 31(2), pp. 199-218.

Nicol, D., Thomson, A and Breslin, C. (2014) 'Rethinking feedback practices in higher education: a peer review perspective', *Assessment and Evaluation in Higher Education*, 39(1) pp.102-122

Norcini, J.J. (2005) 'The Mini Clinical Evaluation Exercise (mini-CEX)' *The Clinical Teacher*, 2(1) pp. 25-30 DOI: 10.1111/j.1743-498X.2005.00060.x

Norcini, J., Brownell, A., Bollela, V., Burch, V., Costa, M.J., Duvivier, R., Galbrith, R., Hays, R., Kent, A., Perrott, V., Roberts, T. (2011) 'Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference' *Medical Teacher*, 33, pp. 206-214

Norman, G., Neville, A., Blake, J.M., Mueller, B (2010) 'Assessment steers learning down the right road: impact of progress testing on licensing examination performance', *Medical Teacher*, 32(6) pp496-9

O'Donovan, B., Rust, C. and Price, M (2016) 'A scholarly approach to solving the feedback dilemma in practice,' *Assessment & Evaluation in Higher Education*, 41(6), pp. 938–949, http://dx.doi.org/10.1080/02602938.2015.1052774

O'Reilly, M., and Parker, N. (2012) 'Unsatisfactory Saturation': a critical exploration of the notion of saturated sample sizes in qualitative research' *Qualitative Research*, 13(2), pp190-197 DOI: https://doi.org/10.1177/1468794112446106

Orsmond, P. and S. Merry (2011) 'Feedback Alignment: Effective and Ineffective Links between Tutors and Students Understanding of Coursework Feedback', *Assessment & Evaluation in Higher Education*, 36 (2), pp. 125-136.

Orsmond, P., Maw, S.J., Park, J.R., Gomez, S., Crook, A.C. (2013) 'Moving feedback forward: theory to practice', *Assessment and Evaluation in Higher Education*, 38(2), pp 240-252

Palinscar, A.S., (1998) 'Social Constructivist perspectives on teaching and learning' *American Review of Psychology*, 49, pp. 345-75

Pallant, J. (2013) *SPSS Survival Manual 5th Edition*, Maidenhead, UK: Open University Press McGraw-Hill Education

Patel, R. Tarrant, C., Bonas, S., Yates, J. and Sandars, J. (2015) 'The struggling student: a thematic analysis from the self-regulated learning perspective', *Medical Education,* 49, pp. 417–426

Patrício, M.F., Julião, M., Fareleira, F, Carneiro, A.V. (2013) 'Is the OSCE a feasible tool to assess competencies in undergraduate medical education?', *Medical Teacher* 35(6), pp.503-514 DOI: 10.3109/0142159X.2013.774330

Pavlov, I.P. (c1927) *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex,* translated and edited by G.V. Anrep, Oxford University Press: London

Pee, B., Woodman, T., Fry, H. & Davenport, E. (2002) 'Appraising and Assessing Reflection in Students' Writing on a Structured Worksheet', *Medical Education*, 36, pp. 575-585.

Pelgrim, E.A.M., Kramer, A.W.M., Mokkink, H.G.A., van der Vleuton, C.P.M. (2013) 'Reflection as a component of formative assessment appears to be instrumental in promoting the use of feedback; an observational study' *Medical Teacher,* 35, pp772-778

Peltier, J.W, Hay, A., Drago, W. (2005) The Reflective Learning Continuum: Reflecting on Reflection, *Journal of Marketing Education*, 27(3), pp. 250-263

Peplow, M (2014) *Nature* 28-8-2014 accessed 25-6-2017 http://www.nature.com/news/social-sciences-suffer-from-severe-publication-bias-1.15787

Perera, J., Mohamadou, G., Kaur, S (2010) 'The use of objective structured self-assessment and peer-feedback (OSSP) for learning communication skills: evaluation using a controlled trial', *Advances in in Health Sciences Education*, 15, pp. 185–193 DOI 10.1007/s10459-009-9191-1

Pett, M.A., Lackey, N.R., Sullivan, J.J. (2003) *Making sense of factor analysis: the use of factor analysis for instrument development in health care research.* Thousand Oaks, Calif: Sage

Pintrich, P. R. (2004) 'A conceptual framework for assessing motivation and self-regulated learning in college students', *Educational Psychology Review,* 16(4), pp. 385-407

Pintrich, P. R., and De Groot, E. V. (1990) 'Motivational and Self-Regulated Learning Components of Classroom Academic Performance' *Journal of Educational Psychology,* 82(1), pp. 33-40

Pintrich, P.R, Smith, D. A.F., Garcia, T., McKeachie, W.J. (1993) 'Reliability and predictive validity of the motivated strategies for learning questionnaire (MSLQ)', *Educational and Psychological Measurement,* 53, pp. 801-813

Pope, C., Zieibland, S. and Mays, N. (2000) 'Qualitative research in health care: Analysing qualitative data,' *BMJ,* 320, 8 Jan 2000, pp.114-116

Price, M., Handley, K., Millar, J., O'Donovan, B. (2010) 'Feedback: all that effort, but what is the effect?', *Assessment & Evaluation in Higher Education*, 35(3), pp. 277–289

Price, M., Handley, K., and Millar, J. (2011) 'Feedback: focusing attention on engagement' *Studies in Higher Education*, 36(8), pp. 879-896

QSR (2012) QSR NVIVO 10, Melbourne, Australia: QSR International Pty Ltd.

Ramani, S. and Krackov, S.K. (2012) 'Twelve tips for giving feedback effectively in the clinical environment', *Medical Teacher*, 34, pp. 787-791

Regehr, G (2010) 'It's NOT rocket science: rethinking our metaphors for research in health professions education', *Medical Education,* 4, pp. 31-39

Reznick RK1, Smee S, Baumber JS, Cohen R, Rothman A, Blackmore D, Bérard M. (1993) 'Guidelines for estimating the real cost of an objective structured clinical examination', *Academic Medicine,* 68(7):513-7.

Richardson, J. T. E., Slater, J. B., Wilson, J. (2007) 'The National Student Survey: development, findings and implications' *Studies in Higher Education,* 32(5) pp557-580 DOI 10.1080/03075070701573757

Ringsted, C., Hodges, B. and Scherpbier, A (2011) 'The research compass': An introduction to research in medical education: AMEE Guide No. 56, *Medical Teacher,* 33(9), pp. 695-709

Robson, C (2011) *Real World Research* 3rd ed. Chichester: John Wiley and Sons Ltd

Rudland, J, Wilkinson, T, Smith-Han, K, Thompson-Fawcett, M. (2008) 'You can do it late at night or in the morning. You can do it at home, I did it with my flatmate. The educational impact of an OSCE', *Medical Teacher,* 30,pp. 206–211

Rudland, J., Wilkinson, T., Wearn, A., Nicol, P., Tunny, T., Owen, C., O'Keefe, M. (2013) 'A student-centred feedback model for educators', *Clinical Teacher,* 10(2) pp.99-102. doi: 10.1111/j.1743-498X.2012.00634.x.

Rush, S., Ooms, A., Marks-Maran, D., Firth, T (2014), Students' perceptions of practice assessment in the skills laboratory: An evaluation study of OSCAs with immediate feedback', *Nurse Education in Practice*, 14, pp. 627-634

Rushforth, H.E (2007) 'Objective structured clinical examination (OSCE): Review of literature and implications for nursing education' *Nurse Education Today*, 27 (5), pp. 481-490 http://dx.doi.org/10.1016/j.nedt.2006.08.009

Rust, C. (2002) 'The impact of assessment on student learning', *Active Learning in Higher Education*, 3(2), pp. 145-158

Rust, C., O'Donovan, B. and Price, M. (2005) 'A social constructivist assessment process model: how the research shows us this could be best practice', *Assessment & Evaluation in Higher Education*, 30(3), pp.231-240

Ryan, R.M., and Deci, E.L. (2000) 'Self-determination theory and the facilitations of intrinsic motivation social development and well-being' *American Psychologist,* 55 (1), pp. 68-78

Sadler, D. R. (1989) 'Formative assessment and the design of instructional system', I*nstructional Science*, 18, pp. 119-144

Sadler, D. R. (1998) 'Formative assessment: revisiting the territory', *Assessment in Education*, 5(1), pp. 77–84.

Saldana, J (2013) *The coding manual for qualitative researchers 2nd ed*. London: SAGE Publications Ltd

Sandars, J. (2009) 'The use of reflection in medical education. AMEE Guide No. 44'. *Medical Teacher*, 31(8), pp. 685-95.

Sandars, J. (2013). 'When I say ... self-regulated learning' *Medical Education,* 47(12), pp. 1162-1163.

Sandars, J. and T. J. Cleary (2011) 'Self-regulation theory: applications to medical education: AMEE Guide No. 58', *Medical Teacher*, 33(11), pp. 875-886.

Sargeant, J. (2008) 'Towards a common understanding of self-assessment', *Journal of Continuing Education in the Health Professions*, 28(1), pp.1-4

Sargeant, J.M, Mann, K.V., Sinclair, D., van der Vleuten, C.P., and Metsemakers, J.F. (2008) 'Understanding the influence of emotions and reflection upon multi-source feedback acceptance and use', *Advances in Health Sciences Education,* 13(3), pp. 275-288.

Sargeant, J.M, Mann, van der Vleuten, C.P., and Metsemakers, J.F. (2009) 'Reflection: a link between receiving and using assessment feedback', *Advances in Health Sciences Education,* 14(3), pp 399-410.

Sargeant, J., Armson, H., Chesluk, B., Dornan, T., Eva, K., Holmboe, E., Lockyer, J., Loney, E., Mann, K., van der Vleuten, C (2010) 'The processes and dimensions of informed self-assessment: a conceptual model', *Academic Medicine*, 85(7), pp.1212-20. doi: 10.1097/ACM.0b013e3181d85a4e.

Schmidts, M.B. (2000) 'OSCE logistics – handheld computers replace checklists and provide automated feedback' *Medical Education*, 34, pp. 947-958

Schön, D. (1991) *The Reflective Practitioner*, Aldershot: Ashgate Publishing Ltd.

Schunk, D.H. (1984) 'Self-efficacy perspective on achievement behaviour' *Educational Psychologist*, 19, pp. 48-58

Schuwirth, L.W.T., van der Vleuten, C.P.M. (2010) How to design a useful test: the principles of assessment, pp195-207 in in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education pp.246-258

Schuwirth, L. W. T & van der Vleuten, C. P. M (2011) 'Programmatic assessment: From assessment of learning to assessment for learning', *Medical Teacher,* 33, pp. 478–485

Shute, V.J. (2008) 'Focus on Formative Feedback' *Review of Educational Research,* 78(1), pp. 153-189 DOI: 10.3102/0034654307313795

Silverman, D. (2000) *Doing qualitative research: A practical handbook*. Thousand Oaks, CA: Sage.

Sinclair, H. K. and Cleland, J. A. (2007) 'Undergraduate medical students: who seeks formative feedback?', *Medical Education*, 41(6), pp. 580-582.

Skinner, B. (1958) 'Reinforcement Today', *American Psychologist*, 13, pp.94-99

Sklar, D. P. (2017) 'Assessment Reconsidered: Finding the Balance Between Patient Safety, Student Ranking, and Feedback for Improved Learning', *Academic Medicine*, 92(6), pp 721-724 doi: 10.1097/ACM.0000000000001687

Sobral, D.T. (2000) 'An appraisal of students' reflection-in-learning' *Medical Education*, 34 (3), pp. 182-187.

Srinivasan, M., Hauer, K.E., Der-Martirosian, C., Wilkes, M., and Gesundheit, N. (2007) 'Does feedback matter? Practice-based learning for medical students after a multi-institutional clinical performance examination', *Medical Education*, 41, pp. 857-865

Streiner, D.L. and Kottner, J. (2014) 'Recommendations for reporting the results of studies of instrument and scale development and testing', *Journal Advanced Nursing* 70(9), pp.1970–1979.

Streiner, D.L. Norman, G.R., and Cairney, J. (2015) *Health Measurement Scales: A Practical Guide to Their Development and Use 5th edn.* Oxford: Oxford University Press

Sundre, D. L. and A. Kitsantas (2004). 'An exploration of the psychology of the examinee: Can examinee self-regulation and test-taking motivation predict consequential and non-consequential test performance?' *Contemporary Educational Psychology,* 29 (1) pp. 6-26.

Swain, S.D., Weathers, D. and Niedrich, R.W. (2008). 'Assessing Three Sources of Misresponse to Reversed Likert Items', *Journal of Marketing Research,* 45(1), pp.116-131.

Swanwick, T. and Buckley, G. (2010) 'Introduction: Understanding Medical Education' in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education pp.xv-xviii

Taras, M. (2005) 'Assessment – summative and formative – some theoretical reflections' *British Journal of Educational Studies,* 53(4) pp 466–478

Telio, S., Regehr, G., Ajjawi, R (2016), 'Feedback and the educational alliance: examining credibility judgements and their consequences', *Medical Education*, 50, pp. 933-942

Thomas, G (2013) *How to do your research project 2nd edition*. London: SAGE Publications Ltd

Thorndike, E. (1911) *Animal Intelligence: Experimental Studies*, New York, USA:Macmillan

Thurlings, M., Vermeulen, M., Bastiaens, T., Stijnen, S (2013) 'Understanding feedback: A learning theory perspective', *Educational Research Review*, 9, pp. 1-15

Tricio, J., Woolford, M. and Escudier, M. (2015) 'Dental students' reflective habits: is there a relation with their academic achievements?', *European Journal of Dental Education*, 19, pp.113–121. doi:10.1111/eje.12111

University of Birmingham (2017) 'Our approach to teaching and learning' available at http://www.birmingham.ac.uk/undergraduate/teaching/approach/index.aspx (accessed 20-5-2017)

Urquhart, L.M., Rees, C.E., Ker, J.S. (2014), 'Making sense of feedback experiences: a muliti-school study of medical students' narratives', *Medical Education*, 48, pp. 189-203

Van der Ridder, J.M., McGaghie, W.C., Stokking, K.M., ten Cate, O.T. (2015) 'Variables that affect the process and outcome of feedback, relevant for medical training: a meta-review', *Medical Education*,49(7), pp.658-73

van der Leeuw, R.M., Slootweg, I.A. (2013) 'Twelve tips for making the best use of feedback', *Medical Teacher*, 35, pp. 348-351

van der Leeuw, R.M. (2014) 'Sharing is caring: dealing with feedback and difficult feelings' *Medical Education*, 48, pp.1034–1042

van der Vleuten and C.P.M., Schuwirth L.W.T. (2005) 'Assessing professional competence: from methods to programmes'. *Medical Education* 39 *(3), pp.* 309–17

van der Vleuten, C.P.M., Schuwirth, L.W.,Driessen, E., Dijkstra, J., Tigelaar D., Baartman, L.K.J., van Tartwijk, J. (2012) 'A model for programmatic assessment fit for purpose', *Medical Teacher ;* 34 *(3), pp.* 205–14

Van Nuland, M., Van den Noortgate, W., van der Vleuten, C., Goedhuys, J. (2012) Optimizing the utility of communication OSCEs: Omit station-specific checklists and provide narrative feedback', *Patient Education and Counseling*, 88, pp. 106-112

Varpio, L., Ajjawi, R., Monrouxe, L. V., O'Brien, B. C. and Rees, C. E. (2017), 'Shedding the cobra effect: problematising thematic emergence, triangulation, saturation and member checking' *Medical Education*, 51, pp. 40–50. doi:10.1111/medu.13124

Veloski, J., Boex, J.R., Grasberger, M.J., Evans, A., Wolfson, D.B. (2006) 'Systematic review of the literature on assessment, feedback and physicians' clinical performance: BEME Guide No. 7'.*Medical Teacher*; 28(2) pp.117-28.

Vygotsky, L. S. (1978).*Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner & E. Souberman., Eds.) (A. R. Luria, M. Lopez-Morillas & M. Cole [with J. V. Wertsch], Trans.) Cambridge, Mass.: Harvard University Press. (Original manuscripts [ca. 1930-1934])

Walford, G (2001) *Doing Qualitative Educational Research.* London: Continuum

Watkins, M.W. (2000) Monte Carlo PCA for parallel analysis (computer software) State College, PA. Ed & Psych Associates.

Watling, C. J. (2014) 'Unfulfilled promise, untapped potential: Feedback at the crossroads', *Medical Teacher*, 36, pp. 692-697

Watling, C., Driessen, E., van der Vleuten, C. P., Vanstone, M. and Lingard, L. (2013) 'Beyond individualism: professional culture and its influence on feedback', *Medical Education*, 47(6) pp. 585-594.

Watling, C.J. and Lingard, L. (2012) 'Toward meaningful evaluation of medical trainees: the influence of participants' perceptions of the process' *Advances in Health Sciences Education*, 17(2), pp. 183–194

Watling, C. (2016) 'The uneasy alliance of assessment and feedback', *Perspectives in Medical Education*, 5, pp. 262-264

Watson, J.B. (1913) 'Psychology as the behaviourist views it', *Psychological Review*, 20(2), pp.158-177

White, C.B. (2007) 'Smoothing Out Transitions: How Pedagogy Influences Medical Students' Achievement of Self-regulated Learning Goals', *Advances in Health Sciences Education*, 12, pp279–297 DOI 10.1007/s10459-006-9000-z

White, C.B., Ross, P.T., Gruppen, L.D. (2009) 'Remediating students failed OSCE performances at one school: the effects of self-assessment, reflection and feedback' *Academic Medicine,* 84(5), pp. 651-654

White, C.B., Ross, P.T., Gruppen, L.D. (2010) 'Self-regulated learning in medical education' in Swanwick, T. (Ed) *Understanding Medical education: Evidence, Theory and Practice.* West Sussex: Wiley-Blackwell and The Association for the Study of Medical Education pp.271-282

White, J.S. and Sharma, N. (2012) 'Who writes what? Using written comments in team-based assessment to better understand medical student performance: a mixed-methods study', *BMC Medical Education* 12:123
DOI:10.1186/1472-6920-12-123

Wiliam (2011) 'What is assessment for learning?' *Studies in Educational Evaluation,* 37, pp. 3-14

Winstone, N., Nash, R., Rowntree, J. & Parker, M. (2016) 'It'd be useful, but I wouldn't use it': barriers to university students' feedback seeking and recipience', *Studies in Higher Education*, pp. 1-16, http://dx.doi.org/10.1080/03075079.2015.1130032

Winstone, N., Nash, R., Parker, M. & Rowntree, J. (2017) Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes', *Educational Psychologist*, 52(1), pp.17–37

Yang, M. and Carless, D. (2013) 'The feedback triangle and the enhancement of dialogic feedback processes', *Teaching in Higher Education*, 18(3), pp. 285-297

Yardley, S., Teunissen, P.W., Dornan, T (2012) 'Experiential learning: AMEE guide No. 63', *Medical Teacher*, 34, pp e102–e115

Zimmerman, B. (1989) 'A Social Cognitive View of Self-Regulated Academic Learning', *Journal of Educational Psychology,* 81(3), pp. 329-339.

Zimmerman, B.J. (2005) 'Attaining Self-regulation: A social cognitive perspective' in Boekaerts, M., Pintrich, P.R., and Zeidner, M. (Eds) *Handbook of Self-Regulation.* St. Louis, MO, USA: Academic Press, Accessed via ProQuest ebrary. Web. 29 October 2015 pp.13-39.

Zimmerman, B.J. (2008), 'Investigating self-regulation and motivation: Historical background, methodological developments and future prospects', *American Educational Research Journal,* 45(1), pp. 166-183

Zimmerman, B.J. (2013) 'From Cognitive modelling to self-regulation: A social cognitive career path', *Educational Psychologist* 48(3), pp135-147

Zimmerman, B.J., Bandura, A., and Martinez-Pons, M. (1992) 'Self-motivation for academic attainment: the role of self-efficacy beliefs and personal goal setting' *American Educational Research Journal,* 29(3), pp. 663-676

APPENDICES

Information for Examiners at RP stations:

- Please familiarise yourself with the candidate info which will advise you of the timings for this question and the role player notes prior to the assessment.
- If the candidate attempts to examine the RP, please stop them and remind them this is a history taking station.
- Please ask the RP (via pointing or signs so that candidates cannot overhear) for their view of how the student performed with regard to communication skills.

Anchor Statements: These statements are designed to help you award an appropriate performance category for each task or question on the individual student's mark sheet. The standard required to score candidate as 'satisfactory' is that of a satisfactory Y4 student at the end of the academic year.

Question / Task	Very poor / Not Done	Unsatisfactory	Satisfactory	Very Good
1. Communication skills Please discuss with RP (once the candidate has left the station via pointing or signs so that candidates cannot overhear).	Does not create a safe non-judgemental environment.	Uses jargon Asks questions but does not listen or makes assumptions Is doctor centred	Shows empathy Listens to patient Is non-judgemental Avoids medical jargon	Student facilitates open, non-judgemental environment Avoids medical jargon Sensitive questioning Shows good listening skills Shows empathy
2. History taking	Does not ask any relevant questions	Lacks focus and has a scattergun approach. Misses significant parts of the history.	Covers most of the urinary symptoms, past history, smoking history and medication history. May not be organised. May not ask about dietary and industrial exposure.	Demonstrates an organised approach to history taking and shows an awareness of the causes of haematuria; asks about dysuria, urinary frequency, previous stones, poor flow, hesitancy, terminal dribble, drug history, eating of beetroot, excessive exercise. Checks smoking history and possible industrial exposure.
3. Eliciting patient's concerns and checks understanding	Does not ask	Assumes the patient is or isn't concerned without checking. Assumes understanding.	Asks the patient if he has any concerns. Checks if patient understands.	Elicits specific concerns about cancer/serious illness

If not already finished, please stop the student at 5 minutes and say: 4. This could be cancer of the urinary tract. Could you tell me what your next steps in management would be?	Does not offer further investigation	Discusses further investigation, but does not appreciate the significance of the clinical findings. Does not specify 2WW referral	Demonstrates an awareness of the significance of the findings and knows this patient requires a 2WW referral to urology	Discusses the need to exclude serious pathology and explains the 2WW referral system clearly. Discussed the value of urine and blood tests whilst you wait for this appointment. Suggests checking the patients BP. May specify excluding infection, checking for protein in the urine, doing FBC, UE, PSA
5. What would you tell the patient to expect when they are seen at the hospital?	No idea	Muddled list of possible investigations, without demonstrating and understanding of the most serious or likely causes of haematuria	Clearly indentifies that the patient is likely to need a cystoscopy to exclude bladder cancer.	Is able to discuss further investigations coherently, explaining why bladder cancer needs to be excluded first and what other investigations may be done if this proves to be negative, e.g. further imaging of the renal tract
If not already covered ask the student "Is there any other advice you should give the patient?" 6. Safety netting	Not considered	Gives unclear instructions	Tells the patient what to do if they do not hear from the hospital within 2 weeks	Gives the patient/describes unambiguous instructions on how they will be contacted, what to do if things don't happen in the expected time frame and how they will get the result of any tests that will be done by the practice.
If the signal for the end of the station has not already sounded, SAY: 'The questions have finished. If you wish to, you may add to any of your answers'				

Instructions for candidate outside the station

(Separate copies are available for the candidate and the examiner inside the station)

This is a Role Played station

This information will also be available for you on the station.

You are a Y4 student at your GP placement and your next patient is Mr Mitchell, who is a 60 years old. You look at his computer record and see that he rarely comes to see the GP and is currently on no regular prescribed medication.

You will have 5 minutes to take a focused history from the patient. In the last 2 minutes the examiner will ask you some questions.

If you complete the history before the examiner asks you to stop, then turn to the examiner and say you have finished.

DO NOT attempt to examine the patient.

The examiner will first take your student ID sticker from you and then you can begin the role play.

Instructions for Role Player

(Separate copies are available for Role player and Examiner)

You are John Mitchell, aged 60, and you have booked an appointment to see a GP, as for the last week your urine has been heavily blood stained. Otherwise you feel absolutely fine and you are passing urine freely with no pain. The blood was still there is morning and you are concerned there is something seriously wrong.

You have collected a urine specimen to show the doctor what it looks like. You collected a sample bottle on arrival at the surgery and did the sample just before coming into your appointment.

- You are generally very well and have had no cause to be concerned about your health prior to this event. You do smoke 15 cigarettes per day, but drink very little alcohol.
- You live with your wife and have three grown up sons.
- You have no family history of cancer.
- You are still working as a secondary school teacher; you love your job and do not want to stop working.
- You do not take any medication.

You want to know what could cause such symptoms and how it will be investigated.

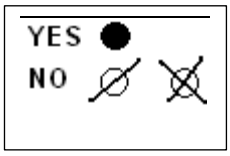
If asked there has been the occasional small clot in your urine. Your urinary stream is as normal, you are not going any more frequently during the day than normal and you only have to get up about once per night which is usual for you, there is no abnormal smell and you have had no dribbling of urine after passing urine. You have not had pain passing urine.

Please start the RP with 'thank you for seeing me doctor. I am really worried as I have been peeing blood for the past week. Here is the sample I did this morning'

Once the candidate has left the station, you will be asked to discuss with the examiner (in whispers or by signs) your view of how the student performed with regard to communication skills.

MB ChB CC3 QRP

Please use a black ball point pen to shade circles completely
Do not put a line or cross through the circle unless you are changing the mark.



Student Details

Label

Affix Student Label Here

If no label, please write student name and number here:

Examiner Details

Label

Affix Examiner Label Here

If no label, please write examiner number here:

FEEDBACK TO STUDENTS
Please provide legible, constructive and specific comments for each student.
This will be scanned and returned directly to the student. Do not write outside the thick line.
Please complete feedback in PEN.

Station: [RP] [Haematuria]
What did you do well?
What could you improve?

TASKS Q[RP]	Very Poor /Not done	Unsatis-factory	Satis-factory	Very Good	Comments on student performance if 'very poor' or 'not done'
1. Communication skills Please discuss with RP (Once the candidate has left the station via pointing or signs so that candidates cannot overhear).	○	○	○	○	
2. History taking	○	○	○	○	
3. Eliciting patient's concerns and checks understanding	○	○	○	○	
AT THE 5 MINUTE SIGNAL STOP THE CANDIDATE (IF NECESSARY) AND SAY: 4. This could be cancer of the urinary tract. Could you tell me what your next steps in management would be?	○	○	○	○	
5. What would you tell the patient to expect when they are seen at the hospital?	○	○	○	○	
IF NOT ALREADY COVERED, ASK THE STUDENT "IS THERE ANY OTHER ADVICE YOU SHOULD GIVE THE PATIENT?" 6. Safety netting	○	○	○	○	

STANDARD SETTING	Below Expected Standard ○	Borderline ○	At Expected Standard ○	Above Expected Standard ○

Professional Behaviour Alert (Only complete if necessary)

Yes ○ **Details:**

Examiner Signature ⇨		Date	

How do students engage with personalised feedback from a summative clinical examination?

Semi-structured Interview Schedule

Before you sat the OSCE
When you were learning during the year or revising for the OSCE did you set yourself any goals about the sort of marks or performance you wanted to achieve?

Directly after the exam
Did you try to estimate what sort of marks or average score you might have achieved?

Did you look up anything after the OSCE exam or speak to other students who sat the same question to find answers before receiving your marks or the written feedback?

When you received the marks from the OSCE
Did your marks match what you thought you would get before and after you sat the exam?

How did you do in the OSCE?

Is this roughly where you usually are in the cohort?

Did you do anything different in your learning or revision this year?

When you read the email that said the individual written feedback comments from the OSCE was available
What did you think/how did you feel?

Did you look at the feedback straight away as soon as you got the email?

If no, why not and when did you look at it (if you did)?

At the time when you read the feedback
How did you feel about the feedback overall?

Did most of it make sense to you generally?

Did the comments generally match with what you thought about your performance on each station? ***(Prior to interview ask for permission to print and bring to interview the actual comments to show and discuss at pertinent points)***

Did what you think about the examiners or questions have an effect on your response to the feedback?

Could you read most of the comments – was the writing legible and the scan dark enough?

Some months have passed now
Did you reflect on aspects of the feedback comments (think about it) and change your mind about what you initially thought about it?

Have you **acted** upon any of the feedback comments? Please explain
If yes what **exactly** did you do? If no, why didn't you (e.g. was it because you didn't understand what the examiner meant or how you could act on it?)

Are you interested in improving on this performance in the next OSCE?

Did you discuss your written feedback with anyone?
(E.g. friend/peer, family member, PM tutor or Y5 clinical academy tutor)?

If you did, did it help you make any more sense of the feedback?
If you didn't, why not? (E.g. you didn't think of it or you didn't have anyone to discuss it with, or didn't think it was worth discussing e.g. didn't value the feedback)?

Do you think it would help students to think about, or make sense or use of the feedback if the Medical School implemented a system to support this for e.g. structured guidance on Canvas or a link to the PM tutor system?

Students who failed at main sit but passed at resit

Link in with questions above to also cover:

What did you think about failing the main sit OSCE?

What did you do after first sit to prepare for the resit OSCE?

After the resit exam
Were there any differences in how you felt about the feedback after the resit OSCE compared to the first sit?

Did you do anything different with the resit feedback comments compared to the first sit ones?

All students
How does this individualised written feedback compare to the other sorts of feedback you get from the OSCE?
Which type of feedback do you prefer / use the most and why?

Remind students that they all received: (print these out and put in folder to show students)

- The score for each of the 12 questions they sat and their own average score for the OSCE overall
- A histogram of the overall average score for all students in the cohort
- The average score for all students for each of the questions in the OSCE
- Generic written comments on students' performance from all examiners who examined on a question, for all the questions used in the examination – not just the ones sat by that student.

Fail students receive in addition:

- Information during a one to one meeting with the Clinical Year Lead which may include a verbal summary of any additional comments written by examiners on each mark sheet relating to specific questions or tasks at that station.

Thank you for talking to me

How do students engage with personalised feedback from a summative clinical examination?

Please use a dark pen to fill the circles completely: ●

Cross through a filled circle if you want to change your answer: ⊗

Please tell me any of the details you would like to **about yourself and your Year 4 OSCE results** by filling in the relevant circles for the 6 questions below:

1. Are you Male ○ Female ○

2. Which course are you on 5 Year course ○ GEC (including Dentists) ○

3. Are you a Home student ○ EEC student ○ Overseas student ○

4. Which age band are you in? 21 or less ○ 22-25 ○ 26-30 ○ 31+ ○

5. What was your overall score in YOUR first sit of the Y4 OSCE (rounded to nearest 1%)?

49 or less ○ 50-54 ○ 55-59 ○ 60-64 ○ 65-69 ○ 70-74 ○ 75-79 ○ 80-100 ○

6. How many stations did you PASS in YOUR first sit of the Y4 OSCE?

1 ○ 2 ○ 3 ○ 4 ○ 5 ○ 6 ○ 7 ○ 8 ○ 9 ○ 10 ○ 11 ○ 12 ○

To help me to improve this survey, if you had any difficulty with understanding or completing any of the questions above, please tell me the question number and the issue(s):

Please continue over the page

SECTION A: Please fill in the appropriate circle to indicate your level of agreement with the following statements about **your actions and thinking on the MB ChB course, overall.**

		Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
1.	I sometimes question the way others do something and try to think of a better way.	○	○	○	○	○
2.	During this course I discovered faults in what I had previously believed to be right.	○	○	○	○	○
3.	I like to think over what I have been doing and consider alternative ways of doing it.	○	○	○	○	○
4.	As a result of this course I have changed the way I look at myself.	○	○	○	○	○
5.	I often reflect on my actions to see whether I could have improved on what I did.	○	○	○	○	○
6.	This course has challenged some of my firmly held ideas.	○	○	○	○	○
7.	I often re-appraise my experience so I can learn from it and improve for my next performance.	○	○	○	○	○
8.	As a result of this course I have changed my normal way of doing things.	○	○	○	○	○

SECTION B: Please fill in the appropriate circle to indicate your level of agreement with statements about **your actions and thinking after sitting the OSCE and whilst waiting for the marks and feedback to be released.**

		Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
9.	I did not think about how I had done on each station because it was too late to change anything.	○	○	○	○	○
10.	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills.	○	○	○	○	○
11.	I asked other students how they had answered questions to help me to work out how I had done on each station.	○	○	○	○	○
12.	I thought to myself about how I had done on each station.	○	○	○	○	○

Please continue over the page

SECTION C: Did you look at the individual feedback comments from your examiners ('What did you do well?' and 'What could you improve?')?

Yes ○ please continue and fill in the appropriate circle to indicate your level of agreement with statements about **your actions and thinking about the individual feedback comments.**

No ○ please let me have any comments on questions 1-12 in the box over the page, thank you.

	Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
13. I looked at all of my individual feedback comments.	○	○	○	○	○
14. I thought about how I could use my individual feedback comments to improve my knowledge or skills.	○	○	○	○	○
15. I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.	○	○	○	○	○
16. I thought about whether I agreed or disagreed with each of my individual feedback comments.	○	○	○	○	○
17. As I wanted to start this year fresh I am not planning to act on the suggested areas to improve in my individual feedback comments.	○	○	○	○	○
18. I read my individual feedback comments as soon as I was able to access them.	○	○	○	○	○
19. I took specific action to improve my knowledge or skills as suggested in my individual feedback comments	○	○	○	○	○
20. Whilst I read my individual feedback comments for interest, I had passed the OSCE, so I did not think about them.	○	○	○	○	○

Please continue over the page

To help me to improve this survey, if you had any difficulty with understanding or completing any of the questions above, please tell me the question number and the issue(s):

Thank you very much for completing this questionnaire

How do students engage with personalised feedback from a summative clinical examination?

Please use a dark pen to fill the circles completely: ●

Cross through a filled circle if you want to change your answer: ●✕

Please write your student registration number in the box below:

By providing your ID number, you are giving your consent for your survey answers to be linked to your basic demographic data and overall OSCE score held by the University. Once linked, your registration number will be immediately deleted from the data set before any analysis to ensure your anonymity. If you would rather not give your ID number, then please go to Q1-6 below to provide this data.

0	○	○	○	○	○	○	○
1	○	○	○	○	○	○	○
2	○	○	○	○	○	○	○
3	○	○	○	○	○	○	○
4	○	○	○	○	○	○	○
5	○	○	○	○	○	○	○
6	○	○	○	○	○	○	○
7	○	○	○	○	○	○	○
8	○	○	○	○	○	○	○
9	○	○	○	○	○	○	○

Please transfer your student number here by completing the relevant circle per column so it can be scanned ⇨

If you did not give your ID number above, please tell me any of the details you would like to about **yourself and your overall OSCE result** by filling in the relevant circles for the 6 questions below:

1. Are you Male ○ Female ○

2. Which course are you on? 5 Year course ○ GEC (including Dentists) ○

3. Are you a UK student ○ EEC student ○ International student ○

4. Which age band are you in? 21 or less ○ 22-25 ○ 26-30 ○ 31+ ○

5. Did you intercalate or take leave of absence (LOA) last academic year? YES ○ NO ○

6. What was your overall score in YOUR first sit of the OSCE (rounded to nearest 1%)?

Write score here ____

Please transfer score here ⇨

10	20	30	40	50	60	70	80	90
○	○	○	○	○	○	○	○	○
1	2	3	4	5	6	7	8	9
○	○	○	○	○	○	○	○	○

Please continue over the page

SECTION A: Please fill in the appropriate circle to indicate your level of agreement with the following statements about **your thinking and actions on the MB ChB course, overall.**

		Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
1.	I sometimes question the way others do something and try to think of a better way.	○	○	○	○	○
2.	During this course I discovered faults in what I had previously believed to be right.	○	○	○	○	○
3.	I like to think over what I have been doing and consider alternative ways of doing it.	○	○	○	○	○
4.	As a result of this course I have changed the way I look at myself.	○	○	○	○	○
5.	I often reflect on my actions to see whether I could have improved on what I did.	○	○	○	○	○
6.	This course has challenged some of my firmly held ideas.	○	○	○	○	○
7.	I often re-appraise my experience so I can learn from it and improve for my next performance.	○	○	○	○	○
8.	As a result of this course I have changed my normal way of doing things.	○	○	○	○	○

SECTION B: Please fill in the appropriate circle to indicate your level of agreement with statements about **your thinking and actions after sitting the OSCE and whilst waiting for the marks and feedback to be released.**

		Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
9.	I thought about the knowledge and skills I was asked to demonstrate at each station.	○	○	○	○	○
10.	I took action to improve my knowledge or skills assessed in the OSCE by looking up information or practising skills.	○	○	○	○	○
11.	I asked other students how they had answered questions to help me to work out how I had done on each station.	○	○	○	○	○
12.	I thought to myself about how I had done on each station.	○	○	○	○	○

Please continue over the page

SECTION C: Did you look at the individual feedback comments from your examiners ('What did you do well?' and 'What could you improve?')?

Yes ○ Please continue and fill in the appropriate circle to indicate your level of agreement with statements about **your thinking and actions about the individual feedback comments.**

No ○ You have now finished the survey, thank you.

	Definitely agree	Agree with reservation	Only to be used if a definite answer is not possible	Disagree with reservation	Definitely disagree
13. I looked at **all** of my individual feedback comments.	○	○	○	○	○
14. I thought about how I could use my individual feedback comments to improve my knowledge or skills.	○	○	○	○	○
15. I plan to take specific action in this academic year to improve my knowledge or skills as suggested in my individual feedback comments.	○	○	○	○	○
16. I thought about whether I agreed or disagreed with each of my individual feedback comments.	○	○	○	○	○
17. I read my individual feedback comments to find out about my knowledge and skills.	○	○	○	○	○
18. I read my individual feedback comments **as soon as** I was able to access them.	○	○	○	○	○
19. I took specific action to improve my knowledge or skills as suggested in my individual feedback comments.	○	○	○	○	○
20. It is important to act on my individual feedback comments to improve my knowledge and skills in this academic year.	○	○	○	○	○

Thank you very much for completing this survey.

Student Participation Information sheet

How do students engage with personalised feedback from a summative clinical examination?

Who is undertaking the research?
Beverley Merricks, Education Development Specialist (Assessment) in the College of Medical and Dental Sciences at the University of Birmingham.

What is the purpose of this research?
This study is intended to garner information to improve my understanding of how students at different levels of performance in a clinical examination engage with the individual feedback comments provided by their examiners.

What information am I being asked to give and why?
If you give your consent to take part in this study, you will be asked to complete a short questionnaire about your propensity to reflect on your studies in general and how you engaged with the individual OSCE feedback comments. You will also be asked to provide either your student ID number so your responses can be linked to your overall OSCE score and some basic demographic information about you, or if you wish to be anonymous at that stage, you will be asked to provide your OSCE score and demographic data directly. If you do give your ID number, the College Assessment Data Analysist will link your survey responses to your OSCE and demographic data and then delete your ID number before passing the encrypted anonymised data to me and deleting his file.

How will the data be stored and who will have access to it?
The primary data from this study will be password protected and held securely on an internal University server. Only I will have direct access to them. Student names or registration numbers will not be stored in the analysis database. The data will be stored for a maximum of ten years after the completion of the research as per current University regulations.

Will the Medical School have access to my responses?
Lecturers or Tutors in the Medical School will not have access to the research data. The findings will be anonymised and it will not be possible for anyone at the Medical School to identify any individual student from them.

Do I have to participate in this study?
No. Participation in the study is completely voluntary.

If I agree to participate can change my mind later?
Yes, if you provided your ID number. Contact Beverley Merricks (via details below) and let her know within 14 days of completing the survey. After this time, your data will be anonymised.

What are the possible risks or disadvantages of taking part?
I cannot foresee any apart from giving up a small amount of your time.

What are the possible benefits of taking part?
This study is intended to provide knowledge about the use students make of the individual OSCE feedback comments. This research may benefit students in the future by informing changes to enable all students to gain the maximum benefit from the feedback provided in order to be able to improve their knowledge and skills and so their future examination performance and potentially their future performance as junior doctors.

Continued over

What will happen to the results of this study?
The results of this study will form part of my Education Doctorate (Ed D) thesis. Relevant findings from the final version of the survey will be reported to the MBChB Curriculum Committees to inform future developments in the way that feedback is provided or students are guided to use the feedback. All findings will be aggregated and no individuals will be identifiable. If you would like to receive (via email) a copy of the aggregated and anonymised results of this study, then please let me know.

Who is supervising the research?
Dr Ian Davison, School of Education, and
Prof Jamie Coleman, Institute of Clinical Sciences,

Who has reviewed the research?
The study has been approved by the Humanities and Social Sciences Ethics Committee at The University of Birmingham (application no ERN_16-0010A).

Any questions?
If you have any questions, please do contact me (details below).

Contact Details:
Ms Beverley Merricks, Education Development Specialist (Assessment)

Medical School
College of Medical and Dental Sciences
University of Birmingham
Vincent Drive
Birmingham
B15 2TT

Consent Form

Name: .....

Please sign below to confirm you have read the information above and have agreed to participate in this study.

Signature: .....

Date: ……..

Thank you very much for agreeing to complete the questionnaire

Missing Data Phase 1 questionnaire		
Category	**Phase 1 Respondent ID Number**	
	Y4	**Y5**
Gender		
Course	4025, 4050	5036
Domicile	4001	5024, 5141
Age		
OSCE score	4011, 4027, 4041, 4042	5012, 5016, 5031, 5137, 5144, 5150, 5151, 5174, 5175, 5181
OSCE stations	4011, 4027, 4041	5012, 5144, 5151,
Q1		
Q2	4049	5024, 5164
Q3		5024
Q4		5176
Q5		
Q6		
Q7		5024
Q8	4059	
Q9	4072	
Q10	4072	
Q11	4072	
Q12	4072	
Q13	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q14	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q15	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q16	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q17	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q18	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056
Q19	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056, 5033
Q20	4004, 4011, 4014, 4015, 4030, 4066, 4072	5013, 5014, 5036, 5042, 5052, 5056, 5033

Missing Data Phase 2 questionnaire			
Category	**Phase 2 Respondent ID Number**		
	Y4	**Y5**	**Graduands**
Gender	None	None	None
Course	**4302, 4309, 4327**	**5169**	None
Domicile	**4321**	None	None
Age	None	None	None
OSCE score	**4174**	None	None
Q1	**4163, 4195**	None	None
Q2	None	None	None
Q3	None	None	None
Q4	**4331**	None	None
Q5	None	None	None
Q6	None	**5210**	None
Q7	None	None	None
Q8	None	None	None
Q9	None	**5192**	None
Q10	None	None	None
Q11	None	None	None
Q12	None	None	None
Q13	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q14	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q15	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q16	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q17	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q18	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q19	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None
Q20	**4136, 4154, 4170, 4188, 4193**	**5143, 5149, 5150, 5191, 5192, 5197, 5205, 5210**	None

NSS Results 2013 and 2016 for Birmingham Medics Course Compared with Full Time Courses in Higher Education Institutions in England

	2013 FT HEI England	2016 FT HEI England	2013 Bham Medicine	2016 Bham Medicine
The teaching on my course				
1 - Staff are good at explaining things.	90	90	92	91
2 - Staff have made the subject interesting.	83	83	91	88
3 - Staff are enthusiastic about what they are teaching.	88	88	91	92
4 - The course is intellectually stimulating.	85	86	97	97
Assessment and feedback				
5 - The criteria used in marking have been clear in advance.	75	77	67	75
6 - Assessment arrangements and marking have been fair.	76	77	70	80
7 - Feedback on my work has been prompt.	68	71	50	58
8 - I have received detailed comments on my work.	70	72	41	52
9 - Feedback on my work has helped me clarify things I did not understand.	65	68	44	50
Academic support				
10 - I have received sufficient advice and support with my studies.	78	80	77	73
11 - I have been able to contact staff when I needed to.	86	87	85	86
12 - Good advice was available when I needed to make study choices.	76	79	73	72
Organisation and management				
13 - The timetable works efficiently as far as my activities are concerned.	80	81	81	74
14 - Any changes in the course or teaching have been communicated effectively.	78	79	64	73
15 - The course is well organised and is running smoothly.	76	77	65	76
Learning resources				
16 - The library resources and services are good enough for my needs.	86	87	91	97
17 - I have been able to access general IT resources when I needed to.	87	89	96	98
18 - I have been able to access specialised equipment, facilities or room when I needed to.	81	83	93	93
Personal development				
19 - The course has helped me present myself with confidence.	81	81	93	93
20 - My communication skills have improved.	84	85	97	97
21 - As a result of the course, I feel confident in tackling unfamiliar problems.	82	83	93	95
Overall satisfaction				
22 - Overall, I am satisfied with the quality of the course.	85	85	90	91

NOTES: All figures in the table represent percentage of respondents who 'definitely' or 'mostly' agreed with the question. NSS Results from http://www.hefce.ac.uk/lt/nss/results/2016 Downloaded 28-1-2017