 Peer-assessment within Dental Education

Tricio-Pesce, Jorge

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International licence. https://creativecommons.org/licenses/by-nc-nd/4.0/

You are free to:
• Share: to copy, distribute and transmit the work

Under the following conditions:
• Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
• Non Commercial: You may not use this work for commercial purposes.
• No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Peer-assessment within Dental Education

Jorge Alberto Tricio-Pesce

A thesis submitted for the degree of

Doctor of Philosophy

King’s College London

2014
Abstract

**Background:** Formative assessment with its feedback component has a powerful effect on students’ learning.

**Aim:** This research aimed to appraise how teaching and assessment practices were organised in dental undergraduate teaching institutions to inform the development and piloting of a novel peer-assessment protocol for undergraduate dental students’ pre-clinical competence and clinical performance. Subsequently, the protocol’s utility as a framework for immediate dialogic peer-feedback to boost students’ academic learning and achievement outcomes as well as their reflective skills, was evaluated.

**Materials and methods:** An initial review of the literature on peer-assessment together with a scrutiny of the King’s College London Dental Institute undergraduate curriculum and assessment practices of 39 selected international dental teaching institutions was undertaken. This underpinned the development of a novel longitudinal, formative and structured peer-assessment protocol based on traditional Workplace-Based Assessment forms to be used as a framework for immediate peer-feedback and self-reflection. Subsequently, the protocol was piloted and later implemented in a larger trained sample to judge its utility towards fostering students’ academic achievements and reflective skills. Thus, following a baseline quantitative reflection skills evaluation, volunteer students assessed their peers’ pre-clinical competence (BDS year-2) and clinical performance (BDS year-5) across the whole academic year. Students’ previous end-of-year examination and baseline reflection skills scores from the study and control groups (those who did and did not exercise the peer-assessment protocol, respectively) were compared to their current end-of-year examination marks and a second reflection skills evaluation score. Students’ feedback narratives and their reasons to participate or not in the peer-assessment protocol, were also analysed.

**Results and Discussion:** Peer-assessment was only used by 19% of the surveyed dental schools. Both pre-clinical and clinical peer-assessment participating students demonstrated a reliable ability to identify those domains where they performed better as well as those which needed improvement. They also detected progress over time. Additionally, students’ peer-assessment scores were positively correlated to their end-of-year examination. Inasmuch as students exercised ten or more peer-assessment encounters, they significantly increased their higher order thinking skills and final examination scores. Peer-feedback narratives from pre-clinical and clinical students differed in their content and sign, but corresponded in their specificity. Previous negative feedback experiences played a notable role in students deciding whether to participate or not.

**Conclusions:** Longitudinal (≥10 encounters), formative and structured peer-assessment and peer-feedback to encourage self-reflection of undergraduate dental students’ pre-clinical and clinical skills, can reliably help them to improve their academic achievement and develop higher order thinking skills.
# Table of Contents

## Volume I

Abstract ........................................................................................................................................... 3
Table of Contents ................................................................................................................................. 4
Table of Figures ................................................................................................................................. 8
Table of Tables ................................................................................................................................ 10
Index of Acronyms ............................................................................................................................ 16
Acknowledgements ............................................................................................................................ 18
General Introduction ......................................................................................................................... 20
Aims .................................................................................................................................................... 22
Chapter 1 .......................................................................................................................................... 25
  Literature Review .............................................................................................................................. 25
  1.1 Introduction ............................................................................................................................... 25
  1.2 Learning theories underpinning peer-assessment and peer-feedback .................................... 26
    Behaviourism ............................................................................................................................... 27
    Cognitivism .................................................................................................................................. 29
    Constructivism ............................................................................................................................ 31
    Experiential learning .................................................................................................................... 33
    Reflective practice learning ........................................................................................................ 35
    Andragogy .................................................................................................................................. 37
  1.3 Learning Theories and the Curriculum ....................................................................................... 40
  1.4 Assessment ............................................................................................................................... 42
    1.4.1 Assessment purposes ......................................................................................................... 43
    1.4.2 Assessment principles ...................................................................................................... 45
    1.4.3 Is there a role for peers in assessment? ............................................................................. 52
    1.4.4 Reported benefits of peer-assessment ............................................................................ 54
    1.4.5 Psychometrics of peer-assessment .................................................................................. 55
    1.4.6 Students’ perception of peer-assessment ......................................................................... 58
    1.4.7 Limitations of peer-assessment ....................................................................................... 61
    1.4.8 Feedback ............................................................................................................................ 62
    1.4.9 Feedback that works ........................................................................................................... 70
    1.4.10 Reflection and learning .................................................................................................... 77
    1.4.11 Clinical workplace-based assessment and the role of peers ......................................... 81
  1.5 Literature review conclusions .................................................................................................... 90
Chapter 2 ........................................................................................................................................... 92
Table of Contents

Teaching and Assessment Approaches in Dental Education ................................. 92
  2.1 Introduction ........................................................................................................ 92
  2.2 Aim ..................................................................................................................... 95
  2.3 Hypotheses ......................................................................................................... 95
  2.4 Materials and Methods ...................................................................................... 95
    2.4.1 KCLDI teaching and assessment ................................................................. 96
    2.4.2 Assessment practices and strategies online survey .................................. 96
  2.5 Results ............................................................................................................... 99
    2.5.1 KCLDI teaching and assessment ................................................................. 99
    2.5.2 Assessment practices and strategies online survey .................................. 103
  2.6 Discussion .......................................................................................................... 105
  2.7 Conclusion ......................................................................................................... 109

Chapter 3 .................................................................................................................. 110

Developing and Piloting a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback ......................................................... 110
  3.1 Introduction ....................................................................................................... 110
  3.2 Aim .................................................................................................................... 111
  3.3 Hypotheses ....................................................................................................... 112
  3.4 Materials and Methods .................................................................................... 112
    3.4.1 Developing the instruments ...................................................................... 112
    3.4.2 Piloting and data collection ...................................................................... 120
    3.4.3 Data analysis .............................................................................................. 122
  3.5 Results ............................................................................................................... 123
  3.6 Discussion .......................................................................................................... 129
  3.7 Conclusion ......................................................................................................... 132

Chapter 4 .................................................................................................................. 134

Utility of a Method to Assess Dental Students’ Reflection Habits ..................... 134
  4.1 Introduction ....................................................................................................... 134
  4.2 Aim .................................................................................................................... 136
  4.3 Hypotheses ....................................................................................................... 136
  4.4 Materials and Methods .................................................................................... 136
    4.4.1 The instrument .......................................................................................... 136
    4.4.2 Participants and data collection ................................................................. 139
    4.4.3 Data analysis .............................................................................................. 139
  4.5 Results ............................................................................................................... 140
    4.5.1 Participants and data collection ................................................................. 140

5
Chapter 6

Implementing a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback

5.1 Introduction ............................................................................................................. 154
5.2 Aim ......................................................................................................................... 156
5.3 Hypotheses .............................................................................................................. 156
5.4 Materials and Methods .......................................................................................... 156
  5.4.1 Instruments for peer-assessment and feedback ................................................. 156
  5.4.2 Instruments to assess academic achievement outcomes ............................... 157
  5.4.3 Instrument for reflection skills assessment ....................................................... 157
  5.4.4 Implementation and data collection ................................................................ 158
  5.4.5 Students’ stimulus ............................................................................................ 160
  5.4.6 Data analysis .................................................................................................... 160
5.5 Results ..................................................................................................................... 164
  5.5.1 Participants ........................................................................................................ 164
  5.5.2 Peer-assessment general data ......................................................................... 164
  5.5.3 Reflection Questionnaires general data ............................................................. 167
  5.5.4 Reliability of peer-DOPS and peer-mCEX ...................................................... 167
  5.5.5 Validity of peer-DOPS and peer-mCEX ......................................................... 168
  5.5.6 Feasibility of peer-DOPS and peer-mCEX ....................................................... 173
  5.5.7 Acceptability of peer-DOPS and peer-mCEX ............................................... 174
  5.5.8 Educational impact .......................................................................................... 175
  5.5.9 Peer-assessment influence on students’ reflective skills ................................. 182
5.6 Discussion ................................................................................................................ 188
5.7 Conclusions ............................................................................................................. 199

Chapter 6

Analysis of Dental Students’ Peer-feedback Narratives from a Prospective Peer-Assessment Protocol

6.1 Introduction ............................................................................................................. 201
6.2 Aim ......................................................................................................................... 202
6.3 Hypotheses ............................................................................................................. 202
6.4 Materials and Methods .......................................................................................... 203
  6.4.1 Quantitative coding of peer-feedback narratives ............................................. 203
### Table of Contents

6.4.2 Qualitative thematic analysis of peer-feedback narratives ........................................ 203  
6.4.3 Focus groups on the reasons peer-assessment participating and non-participating students decided to take part or not in the protocol ........................................ 203  
6.5 Results ........................................................................................................................................ 204  
6.5.1 Quantitative coding of peer-feedback narratives .......................................................... 204  
6.5.2 Qualitative thematic analysis of peer-feedback narratives ........................................... 207  
6.5.3 Focus groups on the reasons peer-assessment participating and non-participating students’ decided to take part or not in the protocol ........................................ 208  
6.6 Discussion .................................................................................................................................... 210  
6.7 Conclusion ...................................................................................................................................... 217  
General Discussion .......................................................................................................................... 219  
Conclusion .......................................................................................................................................... 224  
Future Studies ....................................................................................................................................... 225  
References ........................................................................................................................................... 226

**Volume II**

Appendix Book Contents .................................................................................................................... 267  
Appendix 1 List of Publications and Presentations ............................................................................ 268  
Appendix 2 Dental Students Peer-Assessment: a Prospective Pilot Study.  
European Journal of Dental Education (ID EJE-13-1242.R2 accepted) ........................................ 271  
Appendix 3 Dental Students’ Reflective Habits: Is There a Relation with their  
Academic Achievements? European Journal of Dental Education (ID EJE-14-1382.R1 accepted) .......................................................................................................................... 295  
Appendix 4 Ethical approval ................................................................................................................ 318  
Appendix 5 Mini-Clinical Evaluation Exercise (mini-CEX) ........................................................... 320  
Appendix 6 Direct Observations of Procedural Skill (DOPS) ......................................................... 321  
Appendix 7 Longitudinal Evaluation of Performance (LEP) ............................................................ 322  
Appendix 8 Case-base Discussion (CbD) .......................................................................................... 323  
Appendix 9 Patient Assessment Questionnaire (PAQ) .................................................................... 324  
Appendix 10 Mini-Peer Assessment Tool (mini-PAT) ..................................................................... 325  
Appendix 11 Transcription of the KCLDI syllabus .......................................................................... 326  
Appendix 12 Pre-clinical peer-DOPS for BDS 2 ............................................................................ 495  
Appendix 13 Clinical peer-DOPS for BDS 5 .................................................................................... 497  
Appendix 14 Clinical peer-mCEX for BDS 5 .................................................................................... 499
Table of Figures

Figure A-1  Flowchart of the research chapters with their respective aims and study design.................................................................24

Figure 1-1  Graphical representation of Vygotsky notion of the Zone of Actual Development (ZAD), where the learner can solve problems independently, and the Zone of Proximal Development (ZPD), where the learner can solve problems with support (Vygotsky, 1978)...........................................32

Figure 1-2  Four stages of Kolb’s experiential learning cycle (Kolb, 1976 p. 2) complemented with learning strategies as suggested by Knowles et al. (2012 p. 197).................................................................33

Figure 1-3  Five stages of Schön iterative model of reflective practice following a given action [interpreted from Schön (1983)]........................................35

Figure 1-4  Graphical interpretation of learning theories related to peer-assessment through feedback. .................................................................40

Figure 1-5  Miller’s pyramid for clinical assessment (Miller, 1990) and frequently used assessment methodologies (Kramer et al., 2009).................................46

Figure 1-6  Graphical interpretation of the peer-feedback characteristics that could have a positive or negative influence. This could be used for peer-assessment training purposes. ..................................................................................76

Figure 2-1  Flowchart of the study’s research chapters. The second current chapter is highlighted to help in providing an overall view.........................................92

Figure 2-2  “Preparing for Practice: Dental team learning outcomes for registration” four integrated and mutually supported domains (2012a p. 9).................................93

Figure 2-3  Graphical representation of the percentage of teaching methodologies from all courses together. ..................................................................................99

Figure 2-4  Overall assessment framework within the four levels of Miller’s Pyramid (Miller, 1990)........................................................................100

Figure 2-5  Summative, Formative and Hurdle assessment approaches mapped to the four levels of Miller’s Pyramid (Miller, 1990).................................101

Figure 2-6  Percentage of assessment sessions grouped according to Miller’s pyramid of professional competence, per year.........................................................102

Figure 3-1  Flowchart of the study’s research chapters. The third chapter is highlighted to help in providing an overall view.................................................110

Figure 3-2  Initial header of the peer-assessment DOPS form which maintained the
traditional norm-referenced assessment scale. .......................................................... 113
Figure 3-3 Header of the peer-assessment DOPS form containing now the
criterion-referenced assessment scale. ........................................................................ 114
Figure 3-4 Header of the first option of an educationally-referenced scale with
written explanations of progress presented to students. ........................................... 118
Figure 3-5 Header of the second option of an educationally-referenced scale with
graphical explanations of progress presented to students. ........................................ 119
Figure 3-6 Header of the six-point educationally referenced scale used in both pre-
clinical and clinical peer-assessment instruments which asks the “observing”
student to judge their peer’s ability over time.............................................................. 120
Figure 3-7 Graphical representation of prospective peer-DOPS marks for every
fortnightly assessment occasions (mean and standard deviation). .............................. 126
Figure 4-1 Flowchart of the study’s research chapters. The fourth chapter is
highlighted to help in providing an overall view.......................................................... 134
Figure 4-2 Mean score and ± standard deviation for Habitual Action,
Understanding, Reflection and Critical Reflection scales of the RQ, grouping all
undergraduate students together (UG), clinical postgraduates (CPG) and PhD
trainees.......................................................................................................................... 143
Figure 5-1 Flowchart of the study’s research chapters. The fifth chapter is
highlighted to help in providing an overall view.......................................................... 154
Figure 5-2 Example of a BDS 5 student’s peer-feedback diary designed to help
them recall their peer-feedback and agreed challenges to estimulate self-
reflection (the name of the student has been hidden). ............................................... 159
Figure 5-3 Overall mean peer-DOPS and peer-mCEX scores and standard
deviations by session for BDS 2 and BDS 5 cohorts..................................................... 172
Figure 5-4 Graphical interpretation of Vygotsky ZAD (Zone of Actual
Development) and ZPD (Zone of Proximal Development) (Vygotsky, 1978)
according the results of the present chapter.............................................................. 191
Figure 6-1 Flowchart of the study’s research chapters. The sixth chapter is
highlighted to help in providing an overall view.......................................................... 201
Table of Tables

Table 1-1 Pedagogical and Andragogical models comparison according to Knowles et al. (2012 pp. 298-299). ................................................................. 39

Table 1-2 Percentage of UK NSS higher education full time student’s satisfaction (“mostly agree” and “definitely agree”) to the three questions related to feedback (numbers 7, 8 and 9) at three different years (2007, 2010, 2013), for students from medicine and dentistry, education, as well as the overall score. Source: Higher Education Funding Council for England (2011 p.42) and Unistats (2014). ........................................................................................................... 64

Table 2-1 Taxonomy used to approach the new 2012 GDC Learning Outcomes (General Dental Council, 2012a). ........................................................................ 94

Table 2-2 Template used for mapping the 2012 GDC learning outcomes with BDS teaching and assessing methodologies at KCLDI. .................................................. 96

Table 2-3 Survey designed to assess how dental teaching institutions assessed each of the four GDC “Preparing for Practice” domains. ...................................................... 98

Table 2-4 KCLDI teaching methodologies by course year ........................................ 99

Table 2-5 Learning outcomes range according to the keyword verbs from the proposed taxonomy for each of the four domains ..................................................... 103

Table 2-6 Survey results showing the % of dental teaching institutions using the listed assessment methods for each of the four GDC “Preparing for Practice” domains. ........................................................................................................ 104

Table 3-1 Pre-clinical peer-DOPS checking of the content validity. Each item to be assessed was present in one or more BDS 2 coursebooks learning outcomes. The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93). .................................................................................................................. 115

Table 3-2. Clinical peer-DOPS checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the “Preparing for Practice” GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93). .................................................. 116

Table 3-3. Peer-mCEX checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the “Preparing for Practice” GDC learning outcomes (2012a). The GDC domains have been
coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

Table 3-4 Anonymous questionnaire to study students’ perceptions presented to both BDS 2 and BDS 5 groups at the end of the peer-assessment exercise.

Table 3-5 Variance components estimates for each peer-assessment tools.

Table 3-6 Mean and standard deviation (sd) of the peer-observation and feedback times (minutes), overall peer-assessment score (scale 1 to 6), and students’ perception of the utility of giving and receiving feedback to improve future performance (scale 1 to 6), for each of the studied groups.

Table 3-7 Mean and standard deviation (sd) mark for each of the eleven pre-clinical peer-DOPS items showing the ability of pre-clinical BDS 2 participating students to identify differences (F=3.94, p<0.0001) in their peers’ performance (57 completed forms).

Table 3-8 Mean and standard deviation (sd) mark for each of the thirteen clinical peer-DOPS items showing the ability of clinical BDS 5 participating students to identify differences (F=6.55, p<0.0001) in their peers’ performance (104 completed forms).

Table 3-9 Comparison of the BDS 2 and BDS 5 peer-assessment participating students high stakes end-of-year examination marks with those of their respective classmates who did not take part in the study.

Table 3-10 Selected BDS 2 and BDS 5 students’ feedback comments and agreed challenges and actions, extracted from their peer-DOPS completed forms.

Table 3-11 Anonymous BDS 2 (n=10) and BDS 5 (n=16) students’ perceptions (%) of the peer-assessment protocol after using the peer-DOPS instruments for 6 and 7 occasion, respectively.

Table 4-1 Reflection Questionnaire from Kember et al. (2000) used to assess the reflective habits of students and trainees.

Table 4-2 Number of participants by course of study, mean age and standard deviation (sd) and proportion of students with a previous university degree.

Table 4-3 Mean value for each of the four RQ scales of all subjects and ANOVA with post hoc Tukey’s test results showing inter-scale mean difference and corresponding p-values.

Table 4-4 Mean scores and standard deviations (sd) for each of the four scales of the RQ according to participants’ academic course and study year.

Table 4-5 Separate independent samples t-test results showing number, mean
scores, standard deviations (sd) and p-values for each of the four RQ scales according to participants’ gender, age group and whether they had a previous university degree. .................................................................................................................................144

Table 4-6 Mean value for each of the four RQ scales of all subjects and inter-scale Pearson correlations with the corresponding p-value.................................................................144

Table 4-7 Correlation and corresponding p-value between each RQ scale of BDS year 1-2-3-4-5 subjects and their academic performances in end-of-year examinations. For each examination, the cohort, number of subjects and the examination mean score of the sample as well as that from the whole class ( ) are shown........................................................................................................................................145

Table 4-8 Selection of participants’ comments, study cohort, age, previous degree and their scores in Habitual Action (HA), Understanding (U), Reflection (R), and Critical Reflection (CR)..................................................................................................................................145

Table 5-1 Number of BDS 2 and BDS 5 students according to the number of peer-assessment encounters completed .........................................................................................................................165

Table 5-2 Number of pre-clinical procedures BDS 2 students peer-assessed each other using peer-DOPS according to the perceived complexity assigned to each procedure........................................................................................................................................165

Table 5-3 Number of clinical procedures BDS 5 students peer-assessed each other using peer-DOPS and peer-mCEX according to the perceived complexity assigned to each procedure.................................................................................................................................166

Table 5-4 Summary statistics (mean (sd)) of various measures of BDS 2 and BDS 5 peer-assessment forms........................................................................................................................................166

Table 5-5 D study for different number of encounters for pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX.........................................................................................................................167

Table 5-6 Variance components estimates for each peer-assessment form. ...............168

Table 5-7 The mean and standard deviation (sd) for the 11 peer-DOPS items for pre-clinical peer-assessment of BDS 2 students (N=516 observations)......................................169

Table 5-8 The mean and standard deviation (sd) for the 13 peer-DOPS items for clinical peer-assessment of BDS 5 students (N=576 observations)..................................................169

Table 5-9 The mean and standard deviation (sd) for the 9 peer-mCEX items for clinical peer-assessment of BDS 5 students (N=77 observations)..................................................170

Table 5-10 Anonymous BDS 2 (N=26), BDS 5 (N=43) students’ and Tutors’ (N=9) perceptions (%) of the peer-assessment protocol, assessed at the end of the study. ........................................................................................................................................170
Table 5-11 Correlation between peer-assessment scores and the mean of all end-of-the-year examinations for each of the three tested instruments. ........................................ 171

Table 5-12 Selected anonymous students’ and tutors’ free text comments on the peer-assessment protocol................................................................. 174

Table 5-13 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students’ average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions). ........................................................................................................ 178

Table 5-14 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students’ average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were essays on clinical scenarios, online SAQs (short answer questions), SNQs (short note questions) and online MCQs (multiple choice questions), and the OSCE (Objective Structured Clinical Examination). ........................................................................................................ 179

Table 5-15 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 114 non-participating students (No peer-assessment), for each of the students’ end-of-year examination marks (SAQs (short answer questions), SNQs (short note questions), MCQ (multiple choice questions) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed...................... 180

Table 5-16 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 87 non-participating students (No peer-assessment), for each of the students’ end-of-year examination marks (SAQs=short answer questions, SNQs=short note questions, MCQs=multiple
choice questions, OSCE=Objective Structured Clinical Examination) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. 

Table 5-17 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants’ academic year and whether they participated in the peer-assessment exercise. 

Table 5-18 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants’ academic year and whether they participated in the peer-assessment exercise. 

Table 5-19 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 31 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. 

Table 5-20 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 18 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. 

Table 6-1 Proportional quantitative coding of BDS 2 and BDS 5 peer-feedback narratives according to their content, sign and specificity. 

Table 6-2 Themes identified in BDS 2 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a). 

Table 6-3 Themes identified in BDS 5 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a). 

Table 6-4 Themes identified during the focus groups as the reasons why students who followed the peer-assessment protocol decide to take part in it. Recurrent
themes are shown in bold. .................................................................209

Table 6-5 Themes identified during the focus groups as the reasons why students decided not to take part in the peer-assessment study. Recurrent theme is shown in bold..................................................................................................................209

Table GD-1 Aims and outcome summary for each research chapter of the thesis...220
**Index of Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADEA</td>
<td>American Dental Education Association</td>
</tr>
<tr>
<td>ADEE</td>
<td>Association for Dental Education in Europe</td>
</tr>
<tr>
<td>ADS</td>
<td>Applied Dental Sciences</td>
</tr>
<tr>
<td>AMEE</td>
<td>Association for Medical Education in Europe</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of Variance</td>
</tr>
<tr>
<td>ATSRP</td>
<td>Assessment of Transferable Skills in Removable Prosthodontics</td>
</tr>
<tr>
<td>BDS</td>
<td>Bachelor of Dental Surgery</td>
</tr>
<tr>
<td>BMD</td>
<td>Biomedical Sciences, Dentistry, Medicine and Natural &amp; Mathematical Sciences Ethical Committee</td>
</tr>
<tr>
<td>CAL</td>
<td>Computer Assisted Learning</td>
</tr>
<tr>
<td>CbD</td>
<td>Case-base Discussion</td>
</tr>
<tr>
<td>CFD</td>
<td>Craniofacial Development</td>
</tr>
<tr>
<td>CNR</td>
<td>Case-Note Review</td>
</tr>
<tr>
<td>CODA</td>
<td>American Commission on Dental Accreditation</td>
</tr>
<tr>
<td>COPDEND</td>
<td>Committee of Postgraduate Dental Deans &amp; Directors</td>
</tr>
<tr>
<td>CRE</td>
<td>Clinical Reasoning Examination</td>
</tr>
<tr>
<td>CSCD</td>
<td>Community Special Care Dentistry</td>
</tr>
<tr>
<td>CSR</td>
<td>Chart-Stimulated Recall</td>
</tr>
<tr>
<td>CTT</td>
<td>Classical Test Theory</td>
</tr>
<tr>
<td>DEP</td>
<td>Dental Evaluation of Performance</td>
</tr>
<tr>
<td>D-EP</td>
<td>Dental Evaluation of Performance</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DOPS</td>
<td>Direct Observation of Procedural Skill</td>
</tr>
<tr>
<td>EMQ</td>
<td>Extended Matching Question</td>
</tr>
<tr>
<td>GABA</td>
<td>Gamma-Amino Butyric Acid</td>
</tr>
<tr>
<td>GDC</td>
<td>General Dental Council</td>
</tr>
<tr>
<td>GDP</td>
<td>General Dental Practice</td>
</tr>
<tr>
<td>GIC</td>
<td>Glass Ionomer Cement</td>
</tr>
<tr>
<td>GMC</td>
<td>General Medical Council</td>
</tr>
<tr>
<td>GRAS</td>
<td>Gröningen Reflection Ability Scale</td>
</tr>
<tr>
<td>GT</td>
<td>Generalizability Theory</td>
</tr>
<tr>
<td>HD</td>
<td>High Definition</td>
</tr>
<tr>
<td>IADR</td>
<td>International Association for Dental Research</td>
</tr>
<tr>
<td>ICA</td>
<td>In-Course Assessment</td>
</tr>
<tr>
<td>IOTN</td>
<td>Index of Treatment Needs</td>
</tr>
<tr>
<td>IRT</td>
<td>Item-Response Theory</td>
</tr>
<tr>
<td>ISCP</td>
<td>Intercollegiate Surgical Curriculum Programme</td>
</tr>
<tr>
<td>IV</td>
<td>Intra-Venous</td>
</tr>
<tr>
<td>KCL</td>
<td>King’s College London</td>
</tr>
<tr>
<td>KCLDI</td>
<td>King’s College London Dental Institute</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>LA</td>
<td>Local Anaesthesia</td>
</tr>
<tr>
<td>LEP</td>
<td>Longitudinal Evaluation of Performance</td>
</tr>
<tr>
<td>MCQ</td>
<td>Multiple-Choice Questions</td>
</tr>
<tr>
<td>MCT</td>
<td>Molecules Cells and Tissues</td>
</tr>
<tr>
<td>Mini-CEX</td>
<td>Mini-Clinical Evaluation Exercise</td>
</tr>
<tr>
<td>Mini-PAT</td>
<td>mini-Peer Assessment Tool</td>
</tr>
<tr>
<td>MOS</td>
<td>Minor Oral Surgery</td>
</tr>
<tr>
<td>MPS</td>
<td>Medical Protection Society</td>
</tr>
<tr>
<td>MSF</td>
<td>Multi-Source Feedback</td>
</tr>
<tr>
<td>NICE</td>
<td>National Institute for Health and Care Excellence</td>
</tr>
<tr>
<td>NSS</td>
<td>National Students Survey</td>
</tr>
<tr>
<td>OHP</td>
<td>Oral Health Promotion</td>
</tr>
<tr>
<td>OSCE</td>
<td>Objective Structured Clinical Examination</td>
</tr>
<tr>
<td>OSCT</td>
<td>Observed Structured Clinical Test</td>
</tr>
<tr>
<td>PAP</td>
<td>Peer-Assessment Protocol</td>
</tr>
<tr>
<td>PAQ</td>
<td>Patient Assessment Questionnaire</td>
</tr>
<tr>
<td>PBL</td>
<td>Problem-Based Learning</td>
</tr>
<tr>
<td>PDC</td>
<td>Primary Dental Care</td>
</tr>
<tr>
<td>Peer-DOPS</td>
<td>Peer-Direct Observation of Procedural Skill</td>
</tr>
<tr>
<td>Peer-mCEX</td>
<td>Peer-Mini-Clinical Evaluation Exercise</td>
</tr>
<tr>
<td>PPE</td>
<td>Personal Protective Equipment</td>
</tr>
<tr>
<td>RQ</td>
<td>Reflection Questionnaire</td>
</tr>
<tr>
<td>SAQ</td>
<td>Short Answer Question</td>
</tr>
<tr>
<td>SBA</td>
<td>Single Best Answer</td>
</tr>
<tr>
<td>SCOT</td>
<td>Structured Clinical Operative Test</td>
</tr>
<tr>
<td>SIG</td>
<td>Special Interest Group</td>
</tr>
<tr>
<td>SNQ</td>
<td>Short Note Questions</td>
</tr>
<tr>
<td>SOLO</td>
<td>Structure of the Observed Learning Outcome</td>
</tr>
<tr>
<td>SPRAT</td>
<td>Sheffield Peer Review Assessment Tool</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Sciences</td>
</tr>
<tr>
<td>TCA</td>
<td>Tricarboxylic Acid Cycle (Krebs cycle)</td>
</tr>
<tr>
<td>TEL</td>
<td>Technology Enhanced Learning</td>
</tr>
<tr>
<td>TMA</td>
<td>Tutor Mark Assessment</td>
</tr>
<tr>
<td>TMJ</td>
<td>Temporomandibular Joint</td>
</tr>
<tr>
<td>WPBA</td>
<td>Workplace-Based Assessment</td>
</tr>
<tr>
<td>ZAD</td>
<td>Zone of Actual Development</td>
</tr>
<tr>
<td>ZPD</td>
<td>Zone of Proximal Development</td>
</tr>
</tbody>
</table>
Acknowledgements

These three years of PhD work have been a wonderful and often overwhelming but positive learning experience. What is written in these pages is only part of it. Like in any other academic educational programme, it has involved learning and dealing with many other enriching individual and social learning experiences that have helped me in the endless duty of enhancing the talents and opportunities I have been blessed with. The most important learning experience has been to balance the professional responsibilities of following a challenging PhD with those memorable family life occasions that will never return. It is clear then now, that this has not been a solo race and therefore I am indebted to many. Above all, I thank God for the above mentioned talents, His blessing, wisdom, favour, faithfulness, guidance and protection which otherwise would have made impossible for me to reach this endpoint.

My deepest thanks to:

Professor Mark Woolford, my first supervisor, who from the very first day I visited the Dental Institute, even before starting this academic challenge, trusted me. His initial inspiration became later a permanent and paramount encouragement and support not only for the successful conclusion of the PhD, but also to complement my programme with the best educational training opportunities available in the UK and overseas. I am grateful to him for being always available to listen to me in those early morning talks and especially to help me in professional as well as personal matters. His wise advices to turn towards my family during those difficult times of the PhD are much appreciated.

Dr Michael Escudier, my second supervisor, for his trust and knowledgeable initial advices on starting to write up the thesis from the very beginning which proved correct. His throughout help to critically understand and interpret the sometimes conflicting and entangled results of the current clinical studies, were crucial. Furthermore, his impressive analytical and astute experience on writing scientific papers and on expressing the intended ideas correctly were essential to get the current chapter’s studies speedily accepted for publication.

Professor Agamemnon Grigoriadis, my postgraduate coordinator, for his advices, prompt availability for a relax chat, and his encouraging formative evaluations of progress during the PhD.

Professor Stephen Dunne, Professor Avijit Banerjee and Dr Lyndon Cabot, for trusting my research proposal and allowing and facilitating me to approach years 2 to 5 BDS students and invite them to participate in the current study protocol. Similarly, I am in debt to all
these student’s tutors for relying on me and permitting my presence during every single pre-clinical and clinical session whilst the study was implemented. They are many, but I cannot fail to mention Dr Helena Lewis-Greene, Dr Sofia Zacharaki, Dr Michael Thomas, Dr Lakis Georgiou, Dr George Aristidou, Dr Dipesh Patel, Dr Charles Szentesi, and Dr Paul Duvoisin.

Professor Tim Newton for his help and support on the reflection studies of this thesis.

All King’s dental students who voluntarily decided to take part in this study and enthusiastically followed the research protocol during the whole academic year. There is no way this study would have been possible without their invaluable contribution and appreciated comments and suggestions.

Mr Manoharan Andiappan for his teachings and patience in the sometimes difficult analysis of the huge amount of data.

The Dental Institute Central Office staff members for their always friendly smile and help in all those big and small details that made my PhD experience easier and enjoyable. Mrs Diane Brook and Mrs Margaret Whatley deserve a special line in these acknowledgements.

My friends at my Homeland University of los Andes in Santiago, Chile. They gave me their unconditional support from the very beginning planning of this UK adventure. My deep appreciation and debt of gratitude goes to all the UAndes family and especially to the former Principal Professor Orlando Poblete, former vice-Principal Professor Jaime Arancibia, and to the current Principal Professor José Antonio Guzmán. I am also in debt to my friends and colleagues at the Dental School Council for endorsing this PhD challenge as if it was their own. José Antonio, Francisco and Cristian, I shall never forget your support.

My practice associate, colleague, and life-long friend David Rosenberg without whose support and encouragement, throughout the process, this PhD would not have been possible: I know the practice and patients have been in “good hands”.

Finally, the whole dream of this PhD would have not been possible without the courage of my wife who never doubted to embark the whole family in this venture. Her faith and wise vision of a better future life together guided us all during the pleasant and difficult moments we have spent in the UK. Her unconditional support and sublime care of our boys during the long working months gave me the peace of mind that allowed my dedication and focus on the PhD. To my children, thank you for taking care of your daddy, for helping me to keep good track of the word counts and celebrating the finished chapters, and especially for keeping the house in silence while I was working.

My beloved family, Carolina, Ignacio and Agustin this PhD thesis is dedicated to you…
General Introduction

Traditionally, the model of dental education has been based on the early teaching of the biologic and clinical foundations which underpin all areas of performance, followed by the increasing practice of technical competencies in a laboratory environment and pre-clinical simulation, organised in different parts of the curriculum. Subsequently, once students are shown to be safe, they move from the laboratory into the work environment. That is the clinical environment with real patients, in a model that integrates instructive teaching with real clinical practice. Thus, a varying number of novice learners are guided by an “expert clinical practitioner”, the instructor, in the tradition of the artisan-apprentice training model (Hendricson and Cohen, 1999).

However, this embodies an ever present challenge for clinical instructors as they are required to simultaneously engage in high-quality patient care by ensuring a right diagnosis and treatment, and assess the clinical skills and reasoning of learners (Bowen, 2006). The former depends on factors that most of the time are outside the instructors’ control such as the patient type, multiple clinical situations and time sensitivity (Hoffman and Donaldson, 2004). This setting does not help in identifying students’ specific difficulties (Audétat et al., 2013) and explains why they are seldom observed and provided with feedback during workplace performance (Norcini and Burch, 2007).

Furthermore, the imparted knowledge takes supremacy over the way this knowledge is given (Clarke, 2011), and information passes from the instructor to the novice learner (Harden et al., 1984). In this model, how can we ensure that future professionals, as a competent workforce, keep on learning? (Southgate and van der Vleuten, 2014). If students get dependant on their instructors’ feedback they will hardly develop the needed capacity to assess their own work, and that of others. Consequently, a more student-centred model has been promoted which focuses on the active process of learning (Nicol and Macfarlane-Dick, 2006) and on what students do to achieve their knowledge (Harden and Crosby, 2000), where collaborative and social peer-assessment are encouraged (Chickering and Gamson, 1999) in order to increase the opportunities and frequency of students’ feedback to develop their understanding, learning and reflection (Dochy et al., 1999; Boud and Falchikov, 2007; Biggs and Tang, 2011).

However, as will be discussed in the subsequent chapters of this thesis, both self- and peer-assessment studies in dental education are scarce and those few available are primarily focused on the marking process and not on the learning effects of the
experience. Chapter 1 explores the literature on the learning theories underpinning peer-assessment and peer-feedback. The review goes into greater depth on assessment purposes and principles in addition to the role of peers in it, in order to clarify the potential educational benefits as well as the limitations of peer-assessment reported in different areas of education. Peer-feedback and self-reflection, as the educational components of peer-assessment, are also analysed in the search for an underpinning model that will allow us to develop a peer workplace-based assessment protocol to be implemented in undergraduate programmes.

As a result of the limited evidence of peer-assessment in dental education, I examine the teaching and assessment practices at King’s College London Dental Institute (KCLDI) as a foundation knowledge of teaching and assessment approaches. Subsequently in Chapter 2, I report an international survey of dental teaching institutions’ assessment practices, especially probing the extent of peer-assessment.

Chapter 3 amalgamates both previous chapters’ results to develop and pilot a longitudinal peers’ protocol of observation, scaffolding, assessment and feedback. The consequent challenge, described in Chapter 4, was to probe a valid and reliable method to assess dental students’ reflection. Chapter 5 depicts the pivotal study of implementing the developed and piloted peer-assessment protocol in combination with the reflection evaluation method, in a larger sample throughout the whole academic year to trial its educational impact. Finally, in order to better understand how the protocol could have influenced students’ skills, Chapter 6 analyses students’ written peer-feedback narratives. In the hope of further developing the protocol, this Chapter also describes the motives that drove students to take part or not in the volunteer peer-assessment exercise.
Aims

Adopting aspects of Social Constructivist theory and Reflective Practice Learning as a theoretical framework (Cook et al., 2007), this PhD research aims to appraise how teaching and especially assessment practices are organised in dental undergraduate teaching institutions to inform the development, piloting and utility evaluation of a new peer-assessment protocol of undergraduate dental students’ preclinical-competence and clinical-performance. This framework was used for immediate dialogic peer-feedback in order to attempt to boost students’ academic learning and achievement outcomes as well as their reflective skills.

In order to meet the above primary aim, each chapter of the project contributed with secondary aims (Figure A-1). All these received full ethical approval from King’s College London Biomedical Sciences, Dentistry, Medicine and Natural & Mathematical Sciences Ethical Committee (reference number BDM/11/12-21) (Appendix 4).

Grouped by chapter, secondary aims are:

Chapter 2
- To investigate the different teaching and assessment methodologies used in dental education using King’s College London Dental Institute (KCLDI) as true model.
- To investigate assessment practices and strategies used by selected dental teaching institutions around the world to measure students’ progress particularly in relation to the use of peer-assessment.

Chapter 3
- To develop and pilot a structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students’ skills in all four clinical, communication, professionalism, and management and leadership General Dental Council (GDC) domains, which would be used as an informed framework for the provision of immediate peer-feedback.

Chapter 4
- To assess the usefulness, in terms of reliability and construct validity of a self-reported Reflection Questionnaire (RQ) as a method of assessing dental students’ and postgraduate trainees’ reflection habits.
Chapter 5

- To evaluate the reliability, validity, feasibility, acceptability and educational impact of the same structured protocol (Chapter 3) of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students’ skills as a framework for the provision of immediate peer-feedback. Further, the influence of the peer-assessment protocol on the students’ reflective skills was also examined.

Chapter 6

- To quantitatively and qualitatively analyse students’ written peer-feedback narratives from the peer-assessment exercise, as well the reasons students decided to take part in, or not take part in the peer-assessment protocol.
Figure A-1 Flowchart of the research chapters with their respective aims and study design.
Chapter 1

Literature Review

1.1 Introduction

Learning is the process of bringing together not only knowledge, but emotional and environmental influences and experiences for acquiring, enhancing, or making changes in one's knowledge, skills, values, and views (Ormrod, 2011). It is also about how we perceive and understand the world, about making meaning (Marton and Booth, 1997). Further, the process of learning has been described as being at the heart of life itself (Jarvis, 1987), and one of the most important activities in which humans engage. It is the core of every educational process (Pritchard, 2013), and we as teachers must understand and be aware of this big responsibility.

Consequently, our role as educators must be assumed as a comprehensive, intellectually challenging and facilitated onward process (Oliver et al., 2008; Jahangiri and Mucciolo, 2011) that enables students to learn how to acquire knowledge, how to make the right choices, how to integrate new developments into existing knowledge and how to apply this new knowledge (Kersten, 1997). According to Harden and Laidlaw (2013), most learners in the healthcare professions are very capable and should have little difficulty in achieving the required learning outcomes. However, they will need help, and professional teachers cannot operate using a cookbook approach blindly following guidelines (Harden et al., 1999). Mounting criticism of the quality and efficiency of teaching in higher education (Light et al., 2009), creates the need for tutors to have knowledge about the underlying learning principals as well as the best educational evidence to inform their decision process to facilitate students learning. After all, the aim of teaching is to make student learning possible (Ramsden, 2003), and unfortunately some academics teach students without having much formal knowledge of how students learn (Fry et al., 2009). However, it is understandable, especially for newcomers to the field of education, that different and new pedagogical practices, educational philosophies and conceptual frameworks, can create some initial confusion (Swanwick, 2011).

Accordingly, teaching skills have become increasingly important in view of the new way of creating and using knowledge which focuses on solving problems (Ramsden, 2003), that is, teachers need to help students to develop the ability to employ knowledge
during their practice in such a manner as to devise, choose and apply appropriate
responses to unexpected and sometimes complex problems of the present and the future
(Light *et al.*, 2009; Barber *et al.*, 2013). This is not an easy task especially for busy
clinical tutors. However, as discussed below, encouraging students’ reflection might
facilitate a problem solving approach to practice.

For this reason it was considered important to start this review of the literature
approaching the learning theories which underpin adult learning and especially peer-
assessment and peer-feedback.

### 1.2 Learning theories underpinning peer-assessment and peer-feedback

The nature of learning, what happens when we learn, how it happens and how to
influence the process, has been the matter of study for centuries (Biggs and Tang,
2011). Explanations of the “what” and “how” when we learn, constitute learning or
educational theories. These are based on evidence or long-term successful experience
and work as useful guiding principles to inform practice (Kaufman, 2003), but not as a
mathematical formula (Dennick, 2012). Until recently, psychologists were concern with
developing the “One Grand Theory of Learning” that covered all learning. This notion
is now dead and there are several learning theories that explain how humans learn
(Biggs and Tang, 2011).

It was deemed relevant to this research project to analyse the main learning theories, as
they will be used as a framework to explain the results of the learning interventions
described in the successive chapters.

There are too many specific theories of learning for them all to be reported here.
However, most of the suggested theories in which peer-assessment and feedback as a
learning activity is grounded, fall into three philosophical frameworks (Falchikov and
Goldfinch, 2000). These are cognitivism (information processing, including insight,
memory, perception, and meta-cognition), constructivism (construct meaning from
experience), and andragogy (self-actualisation and continuing personal development)
(Merriam *et al.*, 2007).

Although ‘behaviourism’ (shaping behaviour through rewards) has been strongly
criticised (Chomsky, 1975; Jarvis, 1987), it will be the first theory to be described here,
as feedback, the learning component of peer-assessment, originates in behaviourism (Thorndike, 1911). Further, the intended reflection stimulus of the peer-assessment protocol implemented in this study and subsequently discussed, also suggests the inclusion and discussion of reflective theories of learning (reflection leads to action and then change).

**Behaviourism**

Behaviourism as a theory began with the work of Thorndike on animal intelligence (1911) and those of Watson (1913) in animal behaviour and child education. The theory centres on learning as the acquisition of new behaviour through ‘conditioning’. It disregards how individuals learn, or the cognitive, or mental, activity involved while this is taking place (Pritchard, 2013). Thus, the attention is focused on the rewarding (reinforcement), of a desired behaviour that occurs as a response of a particular stimulus from the surrounding environment (Knowles et al., 2012). There are two types of conditioning that are described as viable explanations of the way animals and humans can be taught to do something (Pritchard, 2013).

First, the “classical conditioning”, where the desired behaviour to be learnt becomes a ‘reflex’ response to a specific stimulus. A well-known example of this is the work of Ivan Pavlov (1927) who conditioned dogs to salivate (reflex) at the sound of a bell (stimulus), as this was an indication that food was imminent (conditioned stimulus) (Merriam et al., 2007; Pritchard, 2013).

Second, the “operant conditioning”, where the desired behaviour to be learnt is reinforced by a reward or a punishment. Again, known examples of this are the studies of probably the most famous behaviourist, the American psychologist Burrhus Frederic Skinner. He studied the response of rats and pigeons to earn a pellet of food (reward) when they pressed a lever. As this reward continued for the repetition of the action, the animal ‘learnt’ that to be fed it must press the lever (Skinner, 1958). Upon the basis of these and other experiments, educationalists have tried to make comparisons between animal and human learning (Wang, 2012).

Therefore, in behaviourism the environment is what shapes students’ behaviours. Consequently, it is the role of the teacher to provide, on more than one occasion, the required environments that prompt the desired behaviours to be learnt (Knowles et al., 2012). Subsequently, the learning programme starts from the student’s initial
knowledge, moving forward in small broken steps which are easy to master (Cross, 1986 p. 232), and, as Skinner emphasised, continually reinforcing success rather than on punishing failure (to keep motivation high), through the provision of immediate feedback (Thorndike, 1911; Skinner, 1954). As a result of the rewards, those students who benefit the most are the less well motivated, the anxious, uncertain and the failing ones. On the other side, bright students can find this programmes unsatisfying and even boring (Pritchard, 2013). While behaviourist positions of learning are sometimes at odds with the current student-centred focus of adult education (Knowles et al., 2012), they are frequently the foundation for one of the largest segments of adult education, such as job and skills training, and self-instructional packages (Cross, 1986; Knowles et al., 2012 p. 125).

In health education, behaviourism is the basis of many objectives and competency-based curricula, and skills development programmes (e.g. tasks under time conditions) (Merriam, 2001; Taylor and Hamdy, 2013). Modelling behaviour through rewards and the influential factor of feedback reinforcement to learn, is common practice and central in health education (Mann, 2011). For example, a behaviourist learning approach would be the “skills and drills” of resuscitation training, which looks at making these ‘behaviours’ automatic (Ker and Bradley, 2011 p. 169).

However, teaching based on behaviourism is usually criticised as it results in learning that promotes standardisation of the outcomes (Taylor and Hamdy, 2013) and rote learning without understanding (Pritchard, 2013). Further, it is posited that animals learn via reflexes and behaviour modification, whilst humans learn through reflection (Wang and King, 2007). The acquisition of these higher order skills such as learning of language, argued Chomsky (1975), are not explained by behaviourist experiments.

Challenging Skinner’s (1954) behaviourist view, Jackson (2009) labels it as a simplistic approach to learning that treats individuals as subjects who are completely dependent on the environment that surrounds them. Further, the behavioural definition of learning, as the product (behaviour modification) of a particular process, is contended by Jarvis (1987), by stating that learning is both a cognitive process and a product. The present thesis will study both the process and product of the peer-assessment protocol, and consequently, the next learning theories will concentrate on the processes that occur in the person who is learning.
Cognitivism

Cognitive learning theories study human thinking to promote learning and started to supersede behavioural psychology in the late 1950s and 1960s, when psychologists and educators began to stress more complex cognitive processes such as thinking, problem solving, language, concept formation and information processing (Ertmer and Newby, 1993). Thus, cognitivism focuses learning not in behaviour but in the mental and psychological processes of the mind, such as perception, insight, information processing (Taylor and Hamdy, 2013), and memory (Regehr and Norman, 1996; Roediger and Karpicke, 2006; Brown et al., 2014).

Cognitivists, such as Piaget (1952) and Bruner (1966) have had great impact on thinking about learning (Knowles et al., 2012). They posit that learners develop new ideas, constructs, hypotheses and decisions based on their interaction with the world and their prior knowledge that exists in their minds, as an internal purposive mental process (Merriam et al., 2007). Thus, the focus is on how information is received, stored and retrieved by the mind (Ertmer and Newby, 1993). Learning is assimilated when the new experience fits into existing structures and adds to the body of examples. In case the experience does not fit the existing structures, learning is accommodated and the structure must be changed in order to incorporate the new knowledge into a cognitive structure that gives meaning and organisation to the knowledge (Ker and Bradley, 2011).

This newly learnt knowledge is then stored in the long-term memory (Kirschner et al., 2006) in the form of schemas which categorise elements of information according to the manner in which they will be used (Sweller et al., 1998). The development of networks between schemas and the use of keywords allow the learners to link previously related experiences to the new one. Therefore, there is an imperative need for the learner to be presented with meaningful teaching experiences that at the same time are linked to pre-existing ones (previous knowledge) (Merriam et al., 2007) as only if the novel information is useful will it trigger particular responses to learn (Regehr and Norman, 1996). This is usually performed using analogies and metaphors to help the learner conceptualise, organise and retain the new information (Ertmer and Newby, 1993).

Thus, cognitive psychology has helped in the understanding of how knowledge is organised and stored in the brain, how memory works, and how learners can be helped to make meaning of their experiences (Merriam et al., 2007). As behaviourism,
Cognitivism emphasises the role of the environment in facilitating learning. Likewise, the role of practice with feedback is also stressed in both approaches: behaviourism uses it for reinforcement; cognitivism for guiding and supporting accurate mental connections (Kirschner et al., 2006). However, in cognitivism the main focus is on the mental activities of the learner, and as such is more appropriate for explaining complex forms of learning (reasoning, problem solving, information processing) than are those of behaviourism (Ertmer and Newby, 1993).

In health education, the concepts of cognitivism have clarified the important processes of clinical reasoning, decision making and problem solving (Kirschner et al., 2006; Merriam et al., 2007; Mann, 2011; Brown et al., 2014). Furthermore, the instruction method of Problem-Based Learning (PBL) is strongly influenced by cognitive psychology (Norman and Schmidt, 1992). Implementing cognitivism, requires the tutor to establish students’ preconceptions in order to present them with the new knowledge in context and at a suitable level of complexity (Knowles et al., 2012). It is also common practice in a cognitivist approach to teaching to create and facilitate access to experiences in order to facilitate learning (Taylor and Hamdy, 2013). Consequently, students are presented with a cognitive conflict in a given experience, drawing attention to the discrepancy between what they expect according to their current knowledge, and the real experience of the event (new knowledge). This is achieved by asking them questions and establishing a dialogic feedback, in which they get to know that the new ‘ideas’ are better than their previous knowledge (Ker and Bradley, 2011). Feedback is then fundamental and acts as a learning catalyst highlighting the gap between students’ actual knowledge and the level they need (Kluger and DeNisi, 1996; Archer, 2010).

Cognitivism as that from Piaget (1952) and Bruner (1966), has been criticised by the overemphasis on cognitive skills at the expense of emotional development; on knowledge attainment ignoring concept formation or invention; and on individual over common motives (Knowles et al., 2012). This very last issue has been also a matter of discussion in the medical education literature. Swanwick (2005) has argued that a cognitivist approach in which the mind is treated as functioning independently of the learners’ social context, is inadequate to explain the full process of clinical learning. This view has been increasingly supported by other healthcare researchers (van der Zwet et al., 2011; Watling et al., 2013) and the present study investigator, especially in the context of clinical workplace learning.
Constructivism

The primary idea of constructivism is that learners actively construct their own knowledge from a concrete experience (Dennick, 2012), in a cumulate nature- and context-bound, that is, based on the activation of what they already know, have experienced and feel, without separating it from the context in which it is used (Kaufman, 2010; Knowles et al., 2012). Further, learners make personal judgments of the knowledge meaning (Merriam et al., 2007) and so about when and how to modify it (Kaufman, 2010). Thus, constructivists believe that we learn best when we actively construct our own understanding from experiences (Pritchard, 2013).

In contrast to cognitivism, the constructivist perspective does not see the teacher as the transmitter of what needs to be acquired, but as a guide who facilitates the construction of the learner’s own new knowledge, based upon experiences and cultural factors (Kaufman, 2010).

Social constructivism (Jarvis, 1987), based on the work of Lev Semyonovich Vygotsky (1978), goes one step further, stressing the importance of social interaction in the learning process (Morris and Blaney, 2011). Thus, understanding and learning from experiences is more successful when the learner engages in dialogue with a more knowledgeable ‘other’ - a teacher, a peer or others in their environment - than when doing it alone (Vygotsky, 1978). The focus is then on the way the student’s community supports learning (Taylor and Hamdy, 2013). It is important to note, however, that a more knowledgeable ‘other’ does not imply someone older (e.g. parent) or in a position of responsibility for learning (e.g. tutor). It is perfectly possible for a peer, friend or even a younger sibling to take this role, especially in casual and informal situations (Pritchard, 2013).

According to Vygotsky (1978), learning would occur by awaking developmental processes that only operate when the learner is interacting with others, stressing the importance of language or ‘shared talk’ (feedback). Thus, through dialogue ideas are considered, shared and developed (Pritchard, 2013).

Further, Vygotsky introduced a fundamental new approach that has impacted on practice over the last 20 years or so, as the novel learning needs to be matched to the learner’s developmental stage through the Zone of Proximal Development (ZPD) (Morris and Blaney, 2011). The ZPD is a theoretical space into which the learner can
move with the support of a more knowledgeable other (Vygotsky, 1978 p. 86; Pritchard, 2013), in contrast to the Zone of Actual Development (ZAD), where they can work unassisted (Figure 1-1). Thus, learning is what takes place in the ZPD, which is just above the understanding level of the learner (the ZAD), where she or he can work but only with guidance, assistance, support and coaching (feedback) (Morris and Blaney, 2011). The term ‘scaffolding’ was later introduced by Wood et al. (1976) to explain this process of help. Subsequently, at the appropriate time, the scaffold is gradually removed to encourage independence (Kaufman, 2010; Ker and Bradley, 2011). A frequently overlooked observation in Vygotsky’s work is his clear distinction between ‘learning’ and ‘development’. Whilst the latter, as the ZPD, is the maturing of the learner’s psychological functions, and not only what students can do with support. Thus, ‘learning’ involves the acquisition of new skills, without changes in the available psychological functions (Vygotsky, 1978 pp. 79-91; Black and Wiliam, 2009).

Figure 1-1 Graphical representation of Vygotsky notion of the Zone of Actual Development (ZAD), where the learner can solve problems independently, and the Zone of Proximal Development (ZPD), where the learner can solve problems with support (Vygotsky, 1978).
In a constructivist approach, effective teaching needs to identify the students’ current competence level (the ZAD) and then offer help and set challenges that are somewhere ahead (the ZPD) of her or his competence (Kaufman, 2010). Despite what social constructivism having helped in providing models to support students’ learning in dental education (Chadwick et al., 2002; Moore and Kain, 2011; Bridges et al., 2014; Postma and White, 2014), it is limited in explaining the complexities of workbased learning (Lave and Wenger, 1991; Wenger, 1998; Morris and Blaney, 2011).

Further, constructivism in general does not provide the mechanism for how learning takes place. Consequently, experiential learning, as developed by David Kolb (1984) will be explored next to explain how experience can be transformed into knowledge, skills and attitudes.

**Experiential learning**

Experiential learning theory is a holistic model of learning that postulates that “knowledge results from the combination of grasping experience and transforming it” (Kolb, 1984 p. 41). Experiential learning is therefore the ability to learn by adapting ideas as a result of experiences (Barley, 2012). For David Kolb, “learning is not so much the acquisition or transmission of content as the interaction between content and experience, whereby each transforms the other” (Knowles et al., 2012 p.195).

![Figure 1-2 Four stages of Kolb’s experiential learning cycle (Kolb, 1976 p. 2) complemented with learning strategies as suggested by Knowles et al. (2012 p. 197).](image-url)
Kolb’s learning cycle (Figure 1-2), stressing the critical importance of experience to learn, helps to explain the gap between students’ understanding something or having just the knowledge, and having the competence to practice it. This might rationalise the strong component of students’ practice at skills laboratories frequently found in dental curriculums. Experiential learning is therefore ‘active’ by definition (Light et al., 2009 p. 55), and probably due to this is the most widespread theory of learning from experience (Fry et al., 2009 p. 15).

The core of the theory conceives learning as a four stage continuous cycle. Immediate concrete experience (i) is the basis for observation and reflection (ii). These observations are assimilated into theory, formation of abstract concepts and generalizations (iii), from which new implications for action can be deduced, testing implications of concepts in new situations (iv), and these implications or hypotheses then serve as guides in acting to create new experiences (Kolb, 1976 p. 2). Despite in practice the learning process starting with an experience, the experiential learning can begin at any of the four stages (Kaufman, 2010). Kolb’s cycle is a practical model for experiential learning practice (Knowles et al., 2012), and though it emphasises individual learning, has been widely accepted in healthcare education (Mann, 2011). Probably the best accepted aspects of Kolb’s theory in medical education are the use of experiences to test new knowledge, the opportunity to provide feedback to change students practices (Kaufman and Mann, 2010), and the increase motivation for learning and reflection that experiential learning produces in students (Ker and Bradley, 2011). It is important to be aware that in practice, students may get stuck somewhere in the cycle, fail to progress or jump about stages (Fry et al., 2009).

Kolb’s experiential learning theory has been criticised for it oversimplification (Ker and Bradley, 2011), its weaker theoretical evidence than other learning theories (Jarvis, 1987), and being grounded only in the individual learner and neglecting the social context of the experience (Bleakley, 2006; Yardley et al., 2012). Further, it has also been questioned for the scarce information it provides about the types of experiences that may promote the students’ engagement in the cycle (Morris and Blaney, 2011 p. 73). However, and despite reflection being a key part of experiential learning (Fry et al., 2009), the need to examine and analyse experiences through the process of reflection, is insufficiently covered in Kolb’s studies (Boud et al., 1985; Jarvis, 1987). Thus for example, students decision making while performing or team learning “concrete experiences”, such as those studied in the present research, would need a broader
underpinning than just a reflection *after* the event. Consequently, a learning theory that offers a deeper analysis of the reflection process will be discussed next.

**Reflective practice learning**

“Reflective practice” is a relatively new phrase that came into use particularly as a result of the work of Donald Schön (Moon, 2013 p. 80). The Schön theory of reflective practice (1983; 1987) is based on the reflection-change model that considers reflection as leading to action and then change. Further, he contends that reflection is a central element of professional thinking as this is how professionals deal with complex and ill-defined problems. So, the importance of Schön’s model lies in that dental and other healthcare practitioners are usually confronted with this kind of unclear clinical case.

![Diagram of Schön's model of reflective practice](image)

*Figure 1-3 Five stages of Schön iterative model of reflective practice following a given action [interpreted from Schön (1983)].*

In a different position to Kolb’s model (Kolb, 1984), where it is required to go through the complete learning cycle in order to further develop one’s own theory, Schön’s reflective practice model implies learning by simply reflecting critically during the
experience, ‘thinking on our feet’ (Schön, 1983 p. 54). In other words, there is no need to make a ‘mistake’ to reflect upon it. It is only required to readjust the theory through a double-loop learning (Figure 1-3). That is, reflecting upon the theory while in action, without going through the entire cycle of reflection after the event (single-loop) (Finger and Asún, 2001 p. 45).

Schön’s model has been described as highly relevant in healthcare education (Taylor and Hamdy, 2013), as it helps to explain what happens when reflective students interact with patients: they realise their limitations when confronted with challenging experiences; they improvise solutions; then later reflect on what happened; and finally consider how to respond in future similar experiences (Yardley et al., 2012).

Thus, as Schön argues, experienced professionals base their practice largely upon ‘tacit knowledge’, that is, somewhat automatic responses grounded on existing mental schemas that enable them to perform efficiently in daily actions. This is termed professionals’ “knowing-in-action” (Schön, 1983), and has been paralleled to riding a bicycle (Shapiro and Talbot, 1991). Occasionally the bicycle skids, which could be the analogy to an unexpected problem or surprise of a professional in practice. At this time, a process of rethinking the “knowing-in-action” is triggered which makes us restructure the strategy, understand phenomena, or ways of framing problems. This gives rise to “on-the-spot experiment and further thinking that affects what we do in the situation at hand”, and probably later in similar ones as well (Shapiro and Talbot, 1991). This rethinking process is known as “reflection-in-action”, which Schön referred also as to ‘professional artistry’ (Schön, 1987 pp. 22, 28-29). Beyond names, the concept of “reflection-in-action” helps us to understand the decisions we take when faced with the above mentioned ill-defined clinical problems.

Further, Light et al. (2009 p. 14) assert that “reflection-in-action” is an ability to employ professional knowledge in complex events and situations, implying the need for knowledge, whilst for Knowles (2012 p. 188) it is the process of reflection while performing, used when existing mental schemas are no longer appropriate, that is, the checking and modifying of practice at the same time. Reflection-in-action requires high cognitive awareness to decision-making and problem-solving in parallel, and as such is less likely to be biased by emotions and subjectivity compared to the automated responses of knowing-in-action (Barley, 2012).

In contrast to experts, novices practice using conscious thoughts and decision making
Barley, 2012), as they lack professional “knowing-in-action” (they are learning to ride the bicycle). Thus, they tend to cling to rules and procedures, which they normally apply in a mechanical way (Finlay, 2008). However, if students base their practice in the automatic responses of knowing-in-action, they will miss opportunities to think about what they are doing (Mamede and Schmidt, 2004).

Consequently, while teaching reflective learning to novices, it is important to instil in them the need to think while practicing and solving problems (thinking on your feet). Thus, by establishing a reciprocal feedback dialogue of reflection-in-action, they learn from the feedback, change and therefore learn effectively (Schön, 1987; Knowles et al., 2012).

However, Schön also proposed the construct of reflection-on-action which occurs later on, after the event (Schön, 1983 p. 278), and provides opportunities to learn from the earlier decision making process (Mamede and Schmidt, 2004). This is a method that both experts and novices could adopt as a reconstructive and conscious mental review, analysis and evaluation of past experiences to determine what may have contributed to the unexpected, and how this situation may affect future similar events (Kaufman and Mann, 2010).

Both reflection-in-action and reflection-on-action are recurring processes that modify pre-existing knowledge directly (Barley, 2012), and where insights and learning from experiences may be incorporated into future “knowing-in-action” (Schön, 1983; Schön, 1987) (Figure 1-3).

Schön’s theory of reflective practice has been criticised for the lack of conceptual clarity surrounding the term “reflective practice” (Ruth, 2013), as it may refer to “a complex array of cognitively and philosophically distinct methods and attitudes” (van Manen, 1995). Further, the emphasis on individual reflection (Sandywell, 2013 pp. 249, 253) and the questionable evidence to distinguish between reflection in- and on-action (Eraut, 1995), have also created disagreement. However, as the very same author asserts, Schön’s book “The Reflective Practitioner” (1983), is the most quoted book on professional expertise.

**Andragogy**

The term andragogy was introduced by the adult educator Malcolm Knowles as “the art and science of helping adults (or, even better, maturing human beings) learn” to
differentiate adult’s learning – andragogy – from children’s learning - pedagogy - (Knowles, 1980), as he considered they learnt differently. So, andragogy emphasises that adults are self-directed individuals and as such expect to take responsibility for their decisions (Merriam et al., 2007).

Knowles presented his model as a set of four ‘assumptions’ (a fifth and a sixth were later included) (Knowles et al., 2012 pp. 138-140), and not as an empirically based theory (Kaufman and Mann, 2010). Despite this, Knowles’ model has produced big controversies. While Norman (1999) argues that “adult learning” (probably avoiding the term ‘andragogy’) is not a theory, Merriam (2001 p. 8) highlights the impact and usefulness of andragogy “for exploring some of the definitional and philosophical issues related to the evolution of adult education as a scientific discipline, and its strengths and weaknesses as a guide to practice”. Again against it, Hartree (1984 p. 203) stated that its popularity can be attributed to “the fact that it makes the kinds of points that adult educators tend to support instinctively for emotional reasons”, that is, Knowles “says what his audience wants to hear”.

Andragogy as a model is related to several educational, social, philosophical and psychological theories that were clustered by Knowles to clarify that adults learn differently and have particular attitudes towards learning (Taylor and Hamdy, 2013). Accordingly, Knowles et al. (2012 p. 140) clearly identify andragogy as being rooted in humanistic perspectives, primarily concerned with the self-actualisation of the individual, and pragmatic philosophy, where knowledge from experience is valued over the one from formal authority. Further, in andragogy, feedback is usually more important than tests and evaluations (Steinert, 2011), and should be delivered in small doses, with the opportunity to obtain more if needed, in order to stimulate the self-assessment process of learning needs (Knowles et al., 2012 pp. 124, 322).

Probably due to an early criticism, enormous debate stating that his pedagogy and andragogy differentiation was artificial (Taylor and Hamdy, 2013), Knowles later modified his model by describing andragogy and pedagogy as a learning continuum (1984). However, his model has played an important part in freeing adult learning from being treated like children in an appropriate way (Yardley et al., 2012).

The changing emphasis, problems and strategies at different times, that underline both approaches as described by Knowles et al. (2012) are summarised in Table 1-1.
Beyond the controversy, andragogy principles can be incorporated into the design of any educational programme to enhance the learner’s receptivity, relevance and engagement (Merriam et al., 2007). Further, an understanding of the assumptions can also influence pacing, meaning and motivation (Steinert, 2011).

The described learning theories (Figure 1-4) have specific characteristics that further define how individuals learn, and have implications for all of us who have a role as teachers/tutors to become professional educators and thus facilitate and maximise students’ learning.

These theories can provide a guiding framework to understand a given teaching approach and better implement an academic programme. Some teachers teach how they have been taught, while others acquire tacit knowledge of pedagogical principles, perhaps by reflecting on their own teaching events. Fortunately, as posited by McLeod et al. (2008), it is today well recognised that content expertise is not enough to succeed as an educator. Thus, it is accepted that understanding the learning process by knowing educational theory, will result in an improved teaching performance (McLeod et al., 2008; Foster and Laurent, 2013).
1.3 Learning Theories and the Curriculum

Learning theories have a reciprocal relationship with curriculum models and their practical effects on educational practice are very powerful (Grant, 2010). The model and structure of the curriculum is a fundamental element of every educational institution and dental schools are no exception (Oliver et al., 2008).

A simple definition for curriculum, as presented by Kern et al. (2009), is “a planned learning experience”. Somewhat more elaborated is the one from The Institute for International Medical Education that defines curriculum as “an educational plan that spells out which goals and objectives should be achieved, which topics should be covered and which methods are to be used for learning, teaching and evaluation” (Wojtczak, 2002). Moving away from a static concept, in the 1970s Stenhouse (1975) stated that the curriculum should be “open to critical scrutiny”. This allows the interpretation of the curriculum as a dynamic educational plan as ideas change. In fact, both learning theory and pedagogical practice, together, are seen today as a never-ending work in progress (Grant, 2010).
In the 1960s a medical curriculum was expected to contain objectives and experiences based on the needs of the society, the philosophy of scientific thinking and the professional characteristics of physicians (Miller, 1961). Later in the 1970s, when behaviourism was predominant, the advice was to define the curriculum in terms of objectives in behavioural terms shaped by the teacher (Simpson, 1972). Subsequently, the 1980s witnessed a move away from the teacher as the transmitter of knowledge, characteristic of cognitivism (Knowles et al., 2012), towards a systems approach (context bound) that emphasised teaching methods with an active participation of students (Newble and Cannon, 2001), more typical of a constructivist perspective (Biggs and Tang, 2011). Attention has moved during the last 30 years from ‘teaching’ towards ‘learning’ (Grant, 2010).

This brings us back to the last part of Wojtczak’s (2002) definition of curriculum. That is, the ‘learning’ and ‘evaluation’ (assessment). As will be discussed below, the focus on learning is today closely related to assessment, and more specifically to formative assessment or ‘for learning’. However, this is not always the case. For example, in the behaviourist model of curriculum, where the teacher is the source of knowledge, assessment takes the role of a ‘test’ at the end of the process for students to show what they have learnt, that is ‘assessment of learning’ (Fish and Coles, 2005). By contrast, in curriculum models where students take an active role in their learning process and teachers turn out to act more as facilitators, as in a constructivist approach, assessment becomes an ongoing process of understanding of the student achievements, informing their next educational needs. Thus, in this model, assessment for learning is a critical part of the teaching and learning process (Fish and Coles, 2005).

The fact that learning is the central purpose in these curricular models, the development of modules or courses can adopt a more student-focused approach. This would mean that the teacher designing a module would start thinking about what the students need to learn, rather than what she or he will be teaching. The key principle is to start by designing the assessment of the module followed by the design of the content of that module (Fry et al., 2009). Further, the assessment should be part of an integrated system of assessment (Schuwirth and van der Vleuten, 2010), which, as described below, would serve both “assessment for learning” and “assessment of learning” purposes (Ramsden, 2003). The evidence suggests that student-centred models, where the purpose of teaching is to support learning, motivate the learner to adopt a deeper approach to learning (Biggs and Tang, 2011). This model has been described as a useful
and efficient approach to curriculum design in dentistry (Manogue and Brown, 2007).

Despite the strong influence learning theories have on curriculum design, there are a wide variety of factors affecting curriculum planning, for example, resources, culture and ethnographies, available learning environments, healthcare service, regulations, and society as a whole (Oliver et al., 2008; Grant, 2010).

1.4 Assessment

Assessment is generally acknowledged as an essential and fundamental part of the education process (Schuwirth and van der Vleuten, 2010), and it is not only important for the student but also for the teacher and course organiser, the accrediting body and the public as a consumer (Harden and Laidlaw, 2012). Assessment provides a window into what students know and ignore, and how they are thinking (Earl, 2012). It affects their lives as their future directions and careers depend on it (Boud and Falchikov, 2007).

The assessment of students learning has been seen as a work overload both to students and tutors, a process that hinders learning, encourages superficiality and conformity, and that it only needs to be done for universities to maintain standards (Knight, 2012). This despite the evidence that started to be available almost twenty years ago (Barr and Tagg, 1995), in that “students’ assessment is at the heart of an integrated approach to student learning”. Further, by then the shift from universities as institutions to “provide instruction” to one that “produce learning”, was already taking hold (Barr and Tagg, 1995). Today, it is widely accepted that assessment, rather than teaching, has a major influence on students’ learning (Boud and Falchikov, 2007). However, the debate about assessment continues and has now moved to issues such as academic standards, preparing students for employment, measuring quality and providing incentives (Boud and Falchikov, 2007).

This is also reflected in the many different roles “assessment” can take (Harlen, 2007), ranging from a certification procedure leading to a pass/fail decision, to assessment as an evaluative or feedback action in education (Manogue et al., 2002; Boud and Falchikov, 2007; Harlen, 2007). As proposed by Schuwirth & van der Vleuten (2010), assessment in competence-based education is “any purported and formal action to obtain information about the competence and performance of a candidate”. In a like
manner, for Fish and Coles (2005 p. 169) assessment is an “educational activity of recognising and recording learners’ achievements and their development within a specific context and in the light of the quality and scope of the education provided for them”.

Thus, assessment is always undertaken with a specific purpose which unfortunately most see just as producing a mark (Moon, 2013), forgetting it also has four main roles: pedagogy (promote learning), measurement (evaluating competence), standardisation (marking to make progress decisions), and certification (fitness to practice) (Fry et al., 2009 p. 134). Moreover, it is often perceived as an afterthought and a burden at the end of the teaching process. Fortunately, a gradual shift is being observed with an increasing attempt to teach what we assess and assess what we want to train for (Crossley and Jolly, 2012). The rationale for this is the widely accepted notion that “assessment drives learning” (Miller, 1990; Wass et al., 2001; Schoonheim-Klein et al., 2006; Manogue et al., 2011; Norcini et al., 2011; Dolmans and Tigelaar, 2012; van der Vleuten et al., 2012). Thus, if teaching and assessment contents comprise an authentic representation and are within the boundaries of the defined competences and learning outcomes (Biggs and Tang, 2011; Tavakol and Dennick, 2011), they will be sending the students the right signal for them to know where, why and what they are working and learning for, directing their learning towards those desirable outcomes (Oliver et al., 2008; Biggs and Tang, 2011; Manogue et al., 2011). Consequently, assessment is a key and integral part of the curriculum and should be seen as inseparably linked to the learning outcomes and teaching methods (Harden and Laidlaw, 2012). As Melnick (2002) put it, “the curriculum instructs teachers; the exam instructs students what to learn”.

**1.4.1 Assessment purposes**

Assessment can either be summative, to measure students’ achievements, or formative, to enhance their learning (Light et al., 2009). However, some scholars argue that this distinction has become blurred as summative and formative purposes can be mixed in one single assessment (Schuwirth and van der Vleuten, 2010; Harden and Laidlaw, 2012). Others, however, have called for caution, as if the same assessment is used for both formative and summative purposes, it creates a conflicting situation for the students: “they are being asked to display and to hide error simultaneously” (Biggs and Tang, 2011 p. 197). However, and despite agreeing with this latter statement, it does not mean students cannot get “formative” feedback from a “summative” assessment.
Summative assessment is often referred to as assessment of learning and is used to confirm what students know for evidence based decision making or certification purposes (e.g. criteria-referenced assessment) (Sadler, 2005). For example, it can be used to rank students, to select who is admitted to a given programme (norm-referenced admission tests), to make judgments about students’ competence progression at a defined level within their programme, or to certify whether students’ are ‘fit for purpose’ when they complete the programme and have achieved the regulating bodies expected standards (Light et al., 2009; Harden and Laidlaw, 2012). Assessment of learning becomes public and it affects students’ futures. Therefore, it is important that these assessment measurements are credible and defensible.

Historically, educational assessment has largely been assessment of learning (Earl, 2012), as it is an essential part of education (Harlen, 2007), and despite some intense criticism, mainly on its impact on students’ learning, the need for certification has secured the “high-stakes” (Knight, 2006) summative assessment position (Boud and Falchikov, 2007 p. 4). Caution has been highlighted for exam-dominated systems, as students’ strategy to pass examinations will become more important than gaining knowledge, leading inevitably to surface learning (Biggs and Tang, 2011).

However, a counter-movement which probably started with the review of classroom formative assessment by Black and William (1998), has seen an emphasises on assessment for learning and how it influences students’ learning (Boud and Falchikov, 2007). Further, contemporary educational philosophies state that assessment for learning should take priority over assessment of learning (Schuwirth and van der Vleuten, 2011b). Indeed, the word ‘assessment’, is derived from the Latin ad sedere or assidere, “to sit beside or with”, and emphasises the importance of the feedback component of assessment (Wiggins, 1993; Manogue et al., 2011).

Formative assessment involves students’ development, improvement and learning (Rolfe and McPherson, 1995; Light et al., 2009), and it is used to find out the level of students’ understanding to provide them with descriptive and informed feedback on their performance progress, to motivate and guide their future learning (Epstein, 2007; McDowell et al., 2010; Manogue et al., 2011; Harden and Laidlaw, 2012).

However and despite students receiving feedback, it is often the case that they do not understand what it means, and thus, cannot take any action until they share the feedback provider conception (Boud and Falchikov, 2007; Bloxham, 2009; Boud and Molloy,
Consequently, recent research on feedback as part of a formative assessment has gone deeper focusing on the use the student gives to the provided feedback. This, the “ultimate goal”, has been coined assessment as learning (Earl, 2012). Accordingly, in addition to receiving external feedback, assessment as learning emphasises this, the feedback, as a process of metacognition for students to critically self-assess and provide their own feedback so that they become their own best assessors fostering their self-regulation process (Earl, 2012). In assessment as learning, feedback is not just a transfer of ideas or a judgement where and what to improve (Bloxham, 2009), it requires the active process of critical self-assessment so that students create their own understanding, making sense and relating these new transferred ideas to previous knowledge, for an action (Sadler, 1989) and effective learning to occur (Black and Wiliam, 1998). To be successful, this process requires the students to be aware of the standards they are aiming for (Earl, 2012), so that they can judge their own performance in relation to those standards. Involving and giving them some responsibility in the assessment process provides them with the opportunity to learn those standards in a given context (Bloxham, 2009). These concepts, in that students should not just be passive receptors of feedback, have also motivated some scholars to call for a re-examination of feedback in higher education (Nicol, 2010; Sadler, 2010). Further, the concept of continuing development at work (once graduated) requires individuals to be creative, to seek and utilise feedback to increase their productivity and be effective practitioners, and to be able to this, universities need to prepare them with high levels of self-regulatory ability (Boud and Molloy, 2013).

Given its importance to the current study, feedback is discussed separately under its own heading below (page 62).

1.4.2 Assessment principles

The design and setting up of a successful assessment system is not easy. There are many different assessment instruments (each with advantages and disadvantages), described in the vast published research literature on medical/dental education (Manogue et al., 2002; Epstein, 2007; Albino et al., 2008; Kramer et al., 2009; Manogue et al., 2011; Norcini et al., 2011). Teachers should understand the outcomes to be assessed and the need for a blend of assessment methodologies (Manogue et al., 2011) as that “perfect assessment is an illusion” (van der Vleuten, 1996). That is, no single methodology can
test all competencies and performances of Miller’s pyramid layers (Figure 1-5) (Miller, 1990; Epstein and Hundert, 2002; Carr, 2006; Chadwick and Holsgrove, 2009; Davies et al., 2009; Schuwirth and van der Vleuten, 2010).

In order to overcome each assessment method weaknesses (Schuwirth and van der Vleuten, 2010), together with psychometric studies highlighting assessors’ subjectivity and students’ case-specificity of clinical performance (Crossley et al., 2002b), during recent years there has been a move towards multiple ‘mini’ test samples across many different assessment formats, such as the Objective Structured Clinical Examination (OSCE) and a wide variety of Workplace-Based Assessment (WPBA) tools (Norcini et al., 1995; Norcini et al., 2003; Prescott-Clements et al., 2008; Schuwirth and van der Vleuten, 2011a; Crossley and Jolly, 2012).

![Figure 1-5 Miller’s pyramid for clinical assessment (Miller, 1990) and frequently used assessment methodologies (Kramer et al., 2009).](image-url)

However, regulating authorities of education and training programmes demand different strengths of assessment methods (General Dental Council, 2012c). Thus, the widely accepted criteria used to evaluate the strengths of a given assessment method (Watson et al., 2014) proposed by van der Vleuten (1996), includes its reliability, validity, cost and feasibility, acceptability, and educational impact. As the peer-assessment protocol developed and implemented in the subsequent research chapters will be judged against these principles, they will be described in the following pages.
Reliability

An assessment method is reliable when its results are reproducible, that is, the extent to which a test yields the same results on repeated trials (Carmines and Zeller, 1979; Schuwirth and van der Vleuten, 2010), and thus its value will reflect the amount of error, both random and systematic, which are intrinsic to any measurement (Streiner and Norman, 2008). This sampling ‘error’ appraisal is particularly important in assessment methods where candidates are assessed by more than one assessor, for example in an OSCE or longitudinal clinical assessments. Thus, the intra-rater and the inter-rater reliability coefficient will estimate the consistency of ratings within the same observer and between them, respectively (Streiner and Norman, 2008). Psychometric studies have concluded that in these cases, multiple examiners across different cases improve inter-rater reliability (Wass et al., 2001).

According to assessment theory, one of the easiest ways to estimate the reliability of, let’s say, a written assessment, is the test-retest analysis in which the same test is given to the same students in two separate occasions. The correlation of both scores on the two administrations will provide the reliability of the test. If the results are exactly the same, the reliability coefficient will be 1.00. However, invariably, the correlation will be less than perfect (Carmines and Zeller, 1979). Thus, reliability coefficients of ≥0.7 and ≥0.8 are generally accepted for low and high-stakes situations, respectively (Beard et al., 2011).

Another option to estimate reliability of an assessment is the parallel test in which, for example, two groups of candidates are assessed using different questions of the same field that are thought to be equivalent in difficulty (Wass et al., 2001). As before, the correlation between both scores will provide the reliability of the test (Schuwirth and van der Vleuten, 2010).

However, both test-retest and the parallel test are not practical. Consequently, most of the well-known reliability tests in Classical Test Theory (CTT), like Cronbach’s alpha (Cronbach, 1951), use a retrospective approach in which the test is randomly divided into two halves and treated as parallel tests. Thus, in CTT the main assumption is that a student has a true ability (known as “true score”), but due to measurement error (known as “error score”), which will always be present, the candidate does not obtain the exact same score even if the same test is used twice. Consequently, the “obtained score = true score ± error score (DeVon et al., 2007). Then, as explained before, the similarities
between these scores are used to mathematically calculate the reliability. The higher the similarity, the lower the “measurement error” (Streiner and Norman, 2008). Further, CTT works best with multiple-choice questions (MCQ) tests since all students take the same questions, as it focuses on the test and its errors (Schuwirth and van der Vleuten, 2010). However, in the presence of several errors influencing the results of a test, like multiple assessor and students being assessed with, for example, different patients or parallel OSCE’s lines, another reliability theory, namely Generalizability Theory (GT), is said to work better (Tavakol and Dennick, 2012).

GT is therefore used when there are variabilities due to multiple examiners (e.g. hawks and doves), clinical scenarios (e.g. different patients or OSCE stations), and the known students’ clinical case-specificity (Eva et al., 1998), where some do better in some cases than their classmates. Generalizability Theory includes all such components and provides an equivalent to reliability called Generalizability Coefficient (Schuwirth and van der Vleuten, 2010). Through sophisticated calculations, GT allows knowing, for example, whether we need more stations or fewer examiners, to reach a certain level of Generalizability. However, and despite a high number of items in a given test taken by a large numbers of students will get better G Coefficient, they cannot be easily extrapolated to a new situation (Schuwirth and van der Vleuten, 2011a).

The third theory concerned with the measurements of reliability is Item-Response Theory (IRT). As opposed to CTT and GT, IRT can be used to identify the behaviour of individual test items and how they interact with individual student abilities (Tavakol and Dennick, 2012), that is, estimate item difficulty independently of students’ ability, and vice versa. Thus, IRT, which is best used with one thousand or more students, can determine if a low score in a given test is the result of particular items of the test being too difficult or the students’ low ability (Tavakol and Dennick, 2013). Further, the reliability of a test consisting of previously used items from an item bank, or students of different abilities, can be calculated before the test is delivered (Schuwirth and van der Vleuten, 2010).

Validity

In a general sense, any measuring device is ‘valid’ if it does what it is intended to do, and does it cleanly without accidentally including other factors (Carmines and Zeller, 1979). In educational terms, validity is the extent to which the competence that the assessment claims to measure is actually being measured (Schuwirth and van der
Vleuten, 2010). This requires the collection of evidence from different approaches, and the focus is not necessarily on scores or items, but rather inferences made from the instrument (Gregory, 2004 pp. 97-98) Therefore, these inferences can be classified and better understood examining some facets of validity that will be studied later in the research chapters. These are content validity, construct validity, and criterion validity (Streiner and Norman, 2008).

Content validity refers to whether the items or questions on a given test accurately reflect the whole testable domain, and therefore it is not determined by its format but by its content (Schuwirth and van der Vleuten, 2010; Schuwirth and van der Vleuten, 2011a). In other words, are items in the test an effective representation of all possible questions that could be derived from the content? (Gregory, 2004), and there are no irrelevant items (Streiner and Norman, 2008 p. 250). For example, an examination on ‘cariology’, would not have content validity if it only comprises questions on, let’s say, “affected dentine”, neglecting “infected dentine” and other ‘contents’. To ensure adequate and relevant coverage of a given test, a matrix, called ‘blueprint’ (Crossley et al., 2002b) is frequently used to efficiently identify the domain(s) of interest, the appropriate assessment methods, and the number of items per domain or category (Schuwirth and van der Vleuten, 2010). However, despite the importance of content validity, it is not sufficient to ensure ‘validity’ of the test, as it does not provide any evidence from the test scores (Messick, 1995; Streiner and Norman, 2008).

To address this issue, the concept of “construct validity” was introduced by Cronbach and Meehl (1955) almost 60 years ago. Construct validity is a back-to-front approach to hypothesis testing (Crossley et al., 2002b). For example, fourth-year dental students’ manual dexterity, which would be the ‘construct’, might be expected to be better than that from first-year dental students, which would be the ‘hypothesis’. Accordingly, a test measuring students manual dexterity applied to both cohorts, would demonstrate “construct validity” if it shows that fourth-year dental students perform better than first-year ones.

Consequently, construct validity indicates the extent to which the scores from the test support the devised hypothesis based on the knowledge of the underlying construct (Norman and Eva, 2010). Subsequently, every time a test is implemented in different situations, let’s say the manual dexterity test in postgraduate trainees, new hypotheses could be derived, and so construct validity would be a continual task (Streiner and
Further on validity testing, criterion validity is the correlation of a test result with some other measure of the attribute being studied, ideally a “gold standard” which is accepted in the field (Streiner and Norman, 2008 p. 254). Accordingly, a high degree of correlation between the “gold standard”, or criterion variable, and the scores on the testing instrument, would be evidence of criterion validity (DeVon et al., 2007). This is sometimes referred to as the “validity coefficient” (Carmines and Zeller, 1979 p. 18; Gregory, 2004 p. 101).

**Cost and Feasibility**

Good assessment is definitely costly, mainly due to staff training and their time spent on planning, taking and marking the assessment (Boursicot et al., 2011). This is especially true “if judgements are any finer grained than pass-fail” (Knight, 2007 p. 77). Unfortunately, economies of scale are difficult to achieve as most assessment costs are directly proportional to student numbers (Gibbs and Simpson, 2004).

However, it should be borne in mind that investing in assessment is investing in teaching and learning (van der Vleuten, 1996), and that expensive assessment can still be good value (Knight, 2007), so far assessment costs do not overtake teaching costs.

Studies reporting assessment methods feasibility in healthcare education do this with different criteria. Accordingly, the systematic review by Donnon et al. (2014) found that within the heterogeneity of Multi-Source Feedback (MSF) instruments evaluated, feasibility was primarily based on the response rate percentages, and rarely included costs and administration concerns. Similarly, other studies have estimated the feasibility of assessment methods by means of completion rate, completion time, faculty time required for review, to score, and to provide feedback, and satisfaction rating (Ram et al., 1999; Kogan et al., 2003; Torre et al., 2007; Mori et al., 2008; Durning et al., 2012; Tolsgaard et al., 2013).

**Acceptability**

Acceptability is another quality required for sound assessment practice (Hays et al., 2002). Even the best assessment method is useless, and it will not survive (van der Vleuten, 1996), if it is not accepted and credible by both teachers and students (Schuwirth and van der Vleuten, 2010; Norcini et al., 2011).
Further and specifically for students, an assessment method will be better accepted when they perceive its educational value and fairness (Norman et al., 1991). This was the case in the study by Escudier et al. (2014) where a big step was undertaken to control cheating in high-stakes e-assessment of dental students (by introducing privacy screen filters), which for 86% (119/138) of students it increased test fairness. Conversely, Tonni and Oliver (2013a) piloted a reflective e-portfolio in six orthodontic postgraduate trainees who highlighted acceptability issues related to it. In line with the literature, the authors acknowledged that the future successful implementation of the new instrument will depend upon the solution of the acceptability issues identified by both students and mentors.

**Educational impact**

As discussed in the previous section (Assessment purposes in page 43), the educational impact of an assessment method or programme is exemplified by the saying “students don’t do what you expect, students do what you inspect” (Schuwirth and van der Vleuten, 2010 p. 198).

Therefore the driving force of assessment should be exploited to achieve the educational objectives (van der Vleuten, 1996). This must consider the principle of “constructive alignment” between the educational objectives and the assessment objectives (Biggs and Tang, 2011). When they are not aligned, the assessment objectives will prevail (van der Vleuten, 1996).

The degree of educational impact of a teaching intervention is increasingly being expressed in the research literature as the “effect size” (Hojat and Xu, 2004). The effect size, known as $d$ (average post-test - average pre-test / spread) (Hattie, 2012 p. 271), has been described as a useful method for comparing the mean results on different measures, or over time, or between groups (Hattie, 2012), independently of the study sample size. The effect size has become increasingly popular as it can be used as a supplement or an alternative to the statistical significance analysis (Fan, 2001).

The average effect size of schooling on overall student achievement, considering more than 100 factors (including student, home, school, teacher, curricula, and teaching domains), has been reported by Hattie and Timperley (2007) to be $d=0.40$. This means that schooling increases the mean on an achievement test by 0.4 of a standard deviation. Further, this value is today used as a “gold standard” on which to judge the effects of
any educational study (Norcini and Burch, 2007). In a later publication by Hattie based on more than 900 meta-analyses of +50,000 research articles with about 240 million students, he suggests that an effect size of $d < 0.20$ would be small, between 0.30 and 0.60 would be medium, and $> 0.60$ would be large (Hattie, 2012 p. 13).

It should be noted that Cohen (1988 pp. 25-27) had previously classified effect size values in the same three categories but with some differences in their values. For Cohen, a small effect size would be $d = 0.20$ (negligible practical importance); medium would be $d = 0.50$ (moderate practical importance); and large would be $d = 0.80$ (crucial practical importance).

1.4.3 Is there a role for peers in assessment?

The *a priori* answer to the above question is yes, provided it is understood in terms of a dialogue, rather than just a substitute for tutor assessment and feedback (Hamer *et al.*, 2014; Nicol *et al.*, 2014). Further on the rationale for this positive answer, it is today well documented that peer-assessment followed by peer-feedback helps students to learn more effectively (Topping, 2005), as the peer experience should not only provide insights into what they know and can do (limiting the encounter to just scoring), but also steer them to improve (Friedlander and Anderson, 2011). Students not only gain a deeper insight into subject matters, but a significant “ability to engage with and take ownership of evaluation criteria, to make informed judgements about the quality of the work of others, to formulate and articulate these judgments in written form and, fundamentally, the ability to evaluate and improve one’s own work based on these processes” (Nicol *et al.*, 2014).

Peer-assessment can be defined in an educational framework as an arrangement that involves observation of students who have attained the same general level of training or expertise and status in order to judge structured tasks or provide global impressions of the amount, level, value, worth, quality or success of their peers’ work (Topping, 1998; Norcini, 2003a; Finn and Garner, 2011). As a result, students are required to provide their peers with grades, feedback or both (Boud and Falchikov, 2007). Indeed, it has been reported that formative peer-assessment can successfully result in the provision of objective feedback (Falchikov and Goldfinch, 2000; Sargeant *et al.*, 2011), which can benefit and enhance the students’ learning process in several ways.

As discussed below, peer involvement in assessment during higher education can aid
learning when it takes place, but it can help to prepare students for independent and autonomous work as well, and by doing so, for a lifetime of learning (Boud and Falchikov, 2007). Further, peers are a key feature of learning in the workplace and in professional practice, as learning with and from peers is the dominant mode of everyday learning (Falchikov, 1995). Furthermore, peer-assessment has the potential to generate profits in the cognitive, social, affective, and professional skills domains (Topping, 1998), and in healthcare it has a long history of identifying successful practitioners (McCormack et al., 2007).

In healthcare education, peers are in an advantageous position to judge one another’s clinical competencies (McCormack et al., 2007), as they have a frequent and close contact in a variety of contexts not always available to faculty (Al Khalifa et al., 2013). This increased exposure (McDonald, 2010) while performing the tasks and procedures that are being learned under real conditions (Shumway and Harden, 2003), allows them to observe, assess and provide feedback (Sargeant et al., 2011) to each other in a less stressful approach (Evans et al., 2007). Furthermore, producing quality feedback as a fundamental graduate skill (Nicol et al., 2014), is an area of practice that is important to develop in university education as students report feeling ill prepared when entering the medical workforce (Burgess et al., 2013).

These potential advantages might explain the reason why peer-assessment has been recommended and encouraged in many publications (Manogue et al., 2002; van der Vleuten and Schuwirth, 2005; Plasschaert et al., 2007; Mattheos et al., 2008; Kramer et al., 2009; Finn and Garner, 2011; Manogue et al., 2011; Tonni and Oliver, 2013b).

In summery, peer-assessment can be beneficial to promote learning if it (Falchikov, 2007 p. 139):

- Is designed as a learning experience.
- Requires learners to take responsibility for their actions.
- Involves providing, seeking and utilising feedback.
- Encourages a reflective approach to learning.
- Requires students to identify and apply standards and criteria.
- Provides some degree of modelling and/or scaffolding.
- Is practised in a variety of contexts.
1.4.4 Reported benefits of peer-assessment

Dochy et al. (1999) reviewed 63 studies and suggested that the use of a combination of new assessment forms such as peer-assessment to inform self-assessment, encourages students to become more responsible and reflective, as the process requires the students to be fair and accurate with the judgments they make regarding their peers. The enhancement of students’ learning by means of reflection was likewise reported by Falchikov (1995). She also found that peer-feedback improved learning by encouraging analysis and diplomatic criticism, and again in agreement with Dochy et al. (1999), she later argued that peer-assessment was a necessary step to develop self-assessment skills (Falchikov, 2007).

The meta-analysis on peer-assessment in higher education conducted by Falchikov and Goldfinch (2000), found that peer-assessment of academic products and processes corresponded more closely to tutor ratings than did professional practice, probably due to students familiarity with the former skills. They recommended conducting peer-assessment in academic settings where students are normally involved, suggesting that peer-assessment could be successful in any discipline area and at any level.

A large analytical review of the literature on UK authorship of innovative assessment methods (from 1996 onwards) across the disciplines, was carried out by Hounsell et al. (2007). Among the more than 400 analysed studies, a number (N=58) were included in the “student involvement in assessment” category and the majority were peer-assessment. The most recurrent advice was the need for preparation and training; making clear the rationale for involving students in assessment; and using pilot studies. Additionally, implementation benefits included aiding the development of personal and lifelong learning skills, helping students overcome unrealistic expectations, facilitating shared understanding and encouraging integration. Likewise, Nulty (2010) found that peer- and self-assessment facilitate greater student involvement in their learning development, and called for a greater use of both methods in the first year of higher education and onwards.

According to Nicol and Macfarlane-Dick (2006), students who have just learned something are frequently better able to explain it to their peers than tutors, as they use a comprehensible language and a more friendly approach. It is also frequently easier to accept criticism from peers than from instructors. Further, peers exchange different perspectives and strategies to solve problems, and construct new knowledge and
meaning through this social interaction (Jarvis, 1987). However, the assessor students also develop new skills as they need to observe and judge their peers’ work in relation to standards, which are then transferred to their own performance.

Similarly, in relation to the benefits to the observing students, Martineau et al. (2013) compared the learning effects of peers observing each other performing a regular physical examination (n=120), to practicing the same procedure alone (N=65). As a result, those students who observed one another performing the examination learnt more than those who practiced alone (p<0.004). The authors concluded that health sciences educational programs may include opportunities for students to learn from their peers through modelling.

Adding to this study, Nicol et al. (2014) evaluated the perception of 82 first-year engineering students on both giving and receiving feedback. A third (27%) stated they learnt from receiving feedback, while the majority (55%) indicated they did learn from both giving and receiving feedback.

1.4.5 Psychometrics of peer-assessment

With regards to peer-assessment scores validity and reliability, the meta-analysis of Falchikov and Goldfinch (2000) found a mean correlation of 0.69 between peer and teacher marks, indicating definite evidence of agreement. Peers’ global judgements with well understood criteria resembled more closely teachers assessments than assessing several individual dimensions, and therefore had a higher validity. The authors suggested the peer-assessing of several dimensions or criteria together with an overall judgement, as the optimum approach. Further, an important aspect to increase validity is the students’ familiarisation with the peer-assessment criteria. This, together with explaining the students the purpose and goals of the peer-assessment exercise have also been recently stressed (Kamp et al., 2013).

Coming back to the Falchikov and Goldfinch meta-analysis (2000), the authors did not find clear validity differences by subject area, but they highlighted that peers in medically related subjects had a lower tendency to agree with tutors’ judgements. Interestingly, this meta-analysis included two studies in dentistry. One on a dental anatomy peer-evaluation (Denehy and Fuller, 1974) and a second one (Jacobs et al., 1975) on self- and peer-evaluation in orthodontics.

The comprehensive review of Topping (1998), found that, in 58% of the studies
(n.18/31) on reliability and validity, peer-assessment was of adequate reliability and validity in a wide variety of applications, while 23% of the studies reported an unacceptably low reliability and validity. Regardless of this, he also found peer assessment to be generally more reliable than self-assessment.

The reliability of undergraduate dental students peer-assessment was evaluated by Satterthwaite and Grey (2008) when 65 pre-clinical dental students’ ivorine teeth preparations were scored by peers and experienced staff assessors. The data showed no significant differences between scores from both groups ($p = 0.531$). Similarly, another study of the same group also on third-year dental students (Taylor et al., 2013), compared staff grades of 78 typodont full gold crown preparations with peer-assessment marks and measurements from a digital scanning device. The results indicated poor levels of agreement between both staff and peer-assessment marks with the digital mechanism. However, once more, similar levels of agreement were seen between experienced assessors and peer-assessment.

Again in dental education but this time in postgraduate maxillofacial surgeon trainees, Evans et al. (2007) studied the criterion validity of peer- and self-assessment scores comparing them with those from trainers. Their results showed that, on average, peer-assessment reflected trainer scores more accurately ($r = 0.83$ for global rating) than self-assessment ($r = 0.55$ for global rating), and that the latter were significantly higher than those given by peers (26.3 versus 28.4, respectively).

Similarly, Sargeant et al. (2008) found in a review article, that accurate self-assessment appeared to be difficult and, for some, even impossible. They also suggested that peer assessment may be more accurate than self-assessment and that Multi-Source Feedback (MSF), a workplace-based assessment frequently used in healthcare education and discussed later in this Chapter (page 85), may be used to inform self-assessment. Thus, this combination of peer-assessment to inform self-assessment has for long been recognised to foster reflection on the student's own learning process and learning activities (Dochy et al., 1999).

A recent systematic review on the reliability and validity of 22 different instruments for students’ peer-assessment in medical education (Speyer et al., 2011), suggested this practice as an effective format for peer-learning. However, the authors found nothing at all or fragmentary or insufficient data on these instruments’ psychometric characteristics. Accordingly, they highlighted that the use of any instrument for
educational purposes can only be justified by its sufficient reliability and validity as well as the discriminative and evaluative purposes of the assessment.

In line with this, and though in a different field of education, Nilson (2003) criticised peer-feedback approaches as having questionable validity, reliability and accuracy and being prone to problems like “bluntness, superficiality, inaccuracies, inconsistencies, and so forth”. She ascribed these problems to the fact that peers are being asked to perform a “demanding process of evaluation” instead of just ‘comprehension’ and ‘analysis’ with detailed and specific guidelines. Further, she argued that under these latter conditions, providing and receiving feedback can be very useful in facilitating student learning, and a valuable “life skill”.

Supporting this observation, Ogden et al. (2000) used a detailed and simple to use 13-task communication and clinical skills checklist at a summative OSCE station where 125 second-year undergraduate medical students were peer-assessed by Final year dental students while performing a mouth examination. Peer scores were then compared to those from experienced staff dentists. The results showed positive correlations between dental students scores and those from staff members.

A study conducted by Lanning et al. (2011), explored the correlation of self-, peer-, near peer- (student of higher course) and staff assessment of second year dental students’ communication skills. Mean scores were for self-assessment 3.86 (sd=0.06); peer-assessment 4.14 (sd=0.04); near peer-assessment 4.07 (sd=0.04); and staff assessment 3.93 (sd=0.10). The highest correlation was observed between peer- and near peer-assessment scores ($r=0.46, p<0.0001$). The authors attribute this to the familiarity of the training programme between near peers. Further, the correlation between peer- and staff assessment was negative ($r=0.08, p=0.707$). It is arguable that the study did not analyse the possibility of a problem in the design of the rating scale as suggested by Hauser & Bowen (2009), or its compounded items (Mackillop et al., 2011b) that might have appeared confusing specially for novice raters.

Although these and other studies present interesting data regarding the validity and reliability of peer-assessment, it is possible that they overemphasized the comparison between peer and tutor awarding marks and most do not consider the exercise as a learning experience. As posited by Liu and Carless (2006), peer-assessment has the potential to support learning, and ‘measurement’ should not take precedent over ‘learning’.
1.4.6 Students’ perception of peer-assessment

Regarding students’ perception of peer-assessment, there are studies reporting opinions both in favour as well as against it. Twenty years ago, Williams (1992) evaluated students’ attitudes towards new approaches to learning and assessment. Thus, 99 first year business students undertook a self- and peer-assessment exercise and their like and dislike responses were analysed. Ninety percent of the students indicated that they enjoyed and saw benefits in both self- and peer-assessment. The latter included comparison of approaches and of standards and exchange of information. However, these novice students also expressed some drawbacks to the exercise. These were the “criticism of friends” and the “arbitrariness in marking”. When asked about improvements, the most frequent answer students provided was “establish firm marking guidelines”. It should be noted that this study did not include a students’ training in the new assessment approaches.

Similarly, Orsmond et al. (1996) also underlined the importance of the marking criteria in biology students’ peer-assessment. However, despite tutors’ and students’ differences in interpreting these criteria, students not only liked carrying out peer-assessment, but also felt the benefits in terms of developing different facets of their learning process. In the same line, third-year geography students expressed their views of a self- and peer-assessment experience as a valuable and enjoyable learning experience which helped them develop skills in independent research, collaboration and communication (Strachan and Wilcox, 1996). In comparison, 96 second-year medical students, who assessed the professional competence of each other, reported that peers identified more strength and weaknesses than they had considered in their own self-assessment, 38% considered these issues to be important enough to become part of their learning plan. Overall, 53% agreed that getting peer-feedback was helpful, while 22% expressed mixed feelings, or disagreement. Sensibly, feedback comments were considered for students to be more helpful than item ratings (Dannefer et al., 2005).

In an exploratory study of students’ perceptions on assessment validity, Sambell et al. (1997) interviewed different subject students (mainly from social science) and complemented this with other sources of evidence (e.g. nature of tasks and assessment products). They reported the common students’ claim that self- and peer-assessment were helping them to develop reflective skills to make judgments about their and other’s work. Most students also perceived the benefits of understanding assessment criteria
and to relate it to their own work. However, they also expressed their worries about passing judgments on their friends. Further, some felt threatened by their perceptions of assessing their peers subjectively or a lack of confidence to act as fair assessors. Once again, the authors call for careful and rigorous preparation and support for students to ensure a close match between what students expect of the peer-assessment exercise and the reality of what they actually experience.

Other kind of students’ negative perceptions were reported by Beaumont et al. (2011) who conducted semi-structured focus-groups to explore the perceptions on peer-assessment of 37 students before entering university. Though some regarded it as constructive and motivational, a much greater proportion expressed bad experiences related to the level of expertise and reliability of peer assessors as well as plagiarism issues.

Sluijsmans et al. (2001), conducted a study where 27 university educational science students were asked to summatively rate their peers after a six-week PBL module. They also completed an evaluation questionnaire where 74% of the students were in favour of implementing peer assessment, 73% felt capable of assessing each other, but only 7% felt comfortable when assessing their peers. Though students doubted the reliability of the method, the Generalizability Coefficient reached a high of 0.944 for 13 ratings, which as presented above (Assessment principles in page 45), is considered optimal for even high stakes examinations (Beard et al., 2011). Further, students expressed the need to make room for feedback, the assessment criteria appeared to be difficult to interpret, and students with no prior experience in peer-assessment felt uncomfortable. These last issues underpinned the authors’ conclusion, similar to the Williams study (1992), in that there is a need for instruction in peer-assessment in order for students to make reliable judgements.

Interestingly, one year later the same group of researchers (Sluijsmans et al., 2002) reported the effects of peer-assessment training on the performance of student teachers. Fifty students were trained in defining performance criteria, giving feedback and writing assessment reports, whilst 43 made up the control group and received no training. Students with training demonstrated higher quality assessment skill. Further, trained students performed significantly better on subsequent tests than students from the control group. Regarding perceptions, students were significantly more satisfied with the re-designed course. The authors concluded that students could be trained in
assessment skills in order to positively increase their performance.

In another aspect of peer-assessment, 42 undergraduate medical students provided and received anonymous feedback about their presentations on anatomy (Gukas et al., 2008). Results showed that most students were positive and felt comfortable receiving and giving feedback, which they perceived as fair, adequate and helpful, and, interestingly, that receiving feedback made them reflect. However, they also expressed that they would be reluctant to give feedback if anonymity was removed.

In order to avoid the reported peer-assessment and feedback social discomfort and associated responsibility, Burgess et al. (2013), investigated students’ perceptions of their ability to provide peer-feedback covering both positive and negative aspects of their performance using the positive critique method (Pendleton et al., 2003). This includes four steps as: a) ask what went well, b) tell what went well, c) ask what could be improved, and d) tell what could be improved. Ninety four fourth-year medical students were trained, and they then observed, assessed and provided feedback to each other using this method over a two year period during formative long-case clinical examinations. The majority of respondents (90%) found the exercise a valuable learning experience, including knowledge and skills development, as well as professionalism attributes. However, despite that they found the positive critique method useful as it allowed them to deliver feedback in a standard and professional manner, a significant number (42%) of respondents did not feel confident in providing negative feedback to their peers.

A recent study of before and after peer-assessment process students’ perception, on 416 students of 11 different subjects in four fields (Planas Lladó et al., 2013), found that students had a positive view towards peer-assessment both before and after its implementation. They perceived it as a motivating and recommended methodology that facilitated the acquisition of learning at different levels.

This was not the case for 52 first-year electrical engineering undergraduate students who before a peer-assessment intervention were not entirely comfortable or confident in their abilities to assess their peers, despite receiving a thorough training in peer-assessment. After the exercise, however, there was a positive shift overall in both attitudes and confidence (Cheng and Warren, 1997).

Among the very few published studies on dental students’ peer-assessment, Larsen and
Jeppe-Jensen (2008) replaced one station of an educational interdisciplinary OSCE by self- and peer-assessment and a later discussion. Sixty-eight third-year dental students who did not receive training on self- or peer-assessment participated in the study. As a result, self- and peer-assessment marks differed widely. Students’ opinions and perceptions regarding the benefit of self- and peer-assessment were not uniform, though the majority found it useful.

A recently published study on dental students’ perception of peer-assessment after grading their peers’ clinical case assignment (Teich et al., 2014), reported that despite participating students (N=55) feeling well prepared for the peer-assessment task, a significant percentage (43.6%) of the group reported that grading the assignment of their peers was not beneficial for their learning process. A possible explanation for these results might be found in the fact that students were not involved in the feedback components of peer-assessment, limiting their role to the measurement (Liu and Carless, 2006).

1.4.7 Limitations of peer-assessment

Despite the overall potential benefits, there remain a number of limitations of peer-assessment. Among the frequently described problems it is not difficult to find reports on “friendship marking”, resulting in overmarking, and “collusive marking”, resulting in a lack of differentiation within groups (Dochy et al., 1999; Evans et al., 2007; Cushing et al., 2011). On these issues, Papinczak et al. (2007) reported a remarkable finding as medical students taking part in a peer-assessment research in Problem-Based Learning, who left prematurely from the study, indicated that friendship marking or lack of honesty was a very important influence on the decision to withdraw. Attitudes among participants included: “I find it difficult to downgrade my peers”; “it is hard to criticize friends”; “no one wants to criticise others in PBL”; “most people are too afraid to honestly mark their peers”.

Other less commonly mentioned grading problems are “decibel marking”, individuals controlling the highest mark, and “parasite marking”, where students fail to contribute but benefit from group marks (Pond and ul-Haq, 1997; Brown and Knight, 1998 p. 59; Dochy et al., 1999). As a way of preventing these problems, the review by Dochy et al. (1999) suggested a combination of peer assessment with self-assessment or co-assessment.
Problems in the design and implementation of the peer-assessment exercise can be detrimental for students to learn (Falchikov and Goldfinch, 2000). Peer-assessment conducted in a traditional norm-reference approach, does not have the power to enable students to benefit from their involvement (Falchikov, 2007). This issue has also been cited in the medical education literature (Norcini, 2003a), as norm-referenced scales used to judge peers’ workplace performance quality, might prove difficult for junior doctors to use.

Similarly, Arnold (2002) suggested that requesting students to peer-assess each other in all clinical performance dimensions may be subject to a “halo effect” (Gregory, 2004 p. 431) as students might not be able to differentiate between peers’ technical knowledge and skills, and professional behaviours. Thus, the author called to limit clinical peer-assessment only to the latter content.

Students do not always accept peer-assessment and feedback as helpful (Beaumont et al., 2011). They have expressed a certain amount of distrust in fellow students’ abilities to peer-assess (Planas Lladó et al., 2013), as well as a reluctance to accept any responsibility for assessing or criticising their friends (Dannefer et al., 2005).

The process of implementation of a peer-assessment method needs to be rigorous in order to alleviate and ideally overcome any students’ concerns (Sambell et al., 1997). This might explain the reason why some teachers have considered the planning and running of such an exercise to be demanding and time-consuming (Hounsell et al., 2007). However, knowing these issues and incorporating students in the planning process (Strachan and Wilcox, 1996), have been suggested to facilitate and favour peer-assessment implementation (Hounsell et al., 2007).

1.4.8 Feedback

At this point, from the above discussion, it is clear that assessment is an integral component of the learning process. Students’ formative assessment results will tell us where they are so that we can support their learning by providing the right feedback on where they should be going (Earl, 2012). Feedback on error then, is a consequence of the assessed performance and is central to students’ learning (Black and Wiliam, 1998; Hattie and Timperley, 2007), as it is inevitable that they will have misconceptions that need to be confronted and eradicated; otherwise, mistakes will probably be repeated (Neher et al., 1992; Biggs and Tang, 2011). However, for a long time there has been a
lack of a commonly accepted meaning of feedback (Ramaprasad, 1983; van de Ridder et al., 2008).

Ramaprasad (1983) defined feedback, in behavioural sciences, as information about the gap between actual performance level and the reference level, which is subsequently used to alter that gap. In this 30-year old definition, he already pointed out, as discussed below, the need for a feedback that is meaningful, understood and that is acted upon.

Latter, Hattie and Timperley (2007), defined feedback as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding. A teacher or parent can provide corrective information, a peer can provide an alternative strategy, a book can provide information to clarify ideas, a parent can provide encouragement, and a learner can look up the answer to evaluate the correctness of a response”. Despite mentioning ‘parents’ as a possible ‘agent’ provider for feedback, which could be interpreted as a definition targeted for school students, it can also be understood that feedback can have different purposes: corrective, alternative, clarifying, and encouraging. As discussed below, this is extremely important as these and other signs, or moderators, will define the usefulness of the feedback.

A recent feedback definition in higher education offered by Boud and Molloy (2013), states that it “is a process whereby learners obtain information about their work in order to appreciate the similarities and differences between the appropriate standard for any given work, and the qualities of the work itself, in order to generate improved work”. Widening its meaning, Evans (2013) includes exchanges occurring beyond the immediate learning context, actively and/or passively sought and/or received and from a range of sources. This broader approach is particularly important for the evaluation of healthcare students and professionals through multisource feedback (Wright et al., 2012), as described in the next section. In medical education, feedback has been defined as “a way in which learners become aware of the gap between their current level of knowledge or skill and the desired goal” (Wood, 2011).

In spite of explaining the concept, not all these definitions mention the need for the feedback receiver to take an action for learning to happen (Sadler, 2010), or the level of students’ engagement (Evans, 2013), or about the quality of the feedback information, particularly its promptness, tone, amount and level of detail, style, clarity, structure and relevance (Wingate, 2010; Nicol et al., 2014). Consequently, once again, feedback today is not just telling the students where and what to improve (Bloxham, 2009); they
must play an active role during (Carless, 2006) and after the process (Nicol, 2010). Further, as presented by Boud and Molloy (2013), the current duty of feedback is not just to improve performance ‘now’, but also on the capacity of the learner to better manage future and different tasks.

This reconceptualization of feedback in higher education, has come as a response to the learners’ complaint that they never receive feedback (Branch and Paranjape, 2002) or that it is not enough (Boud and Molloy, 2013), which has consistently been reflected in the low levels of students’ satisfaction with feedback in the National Students Survey (NSS) in the UK (Higher Education Funding Council for England, 2011 pp. 8-14). The NSS is undertaken by final year undergraduate higher education students in the UK, and is recognised as an important measure of student satisfaction (National Student Survey, 2014). Results from the NSS are publicly available (Unistats, 2014), allowing comparison of different universities and courses. Thus, it is within every university’s best interests to achieve the highest NSS scores as possible (Holmes, 2014).

Accordingly, an analysis of the report of the findings and trends of the sixth annual NSS carried out in 2010, published by the Higher Education Funding Council for England (2011 p.42), as well as the available data from Unistats (Unistats, 2014), allowed comparison of the level of students’ satisfaction in those three questions related to feedback for medical and dental students (as both subjects are presented together), those following education studies, and the global score (Table 1-2).

<table>
<thead>
<tr>
<th>Question</th>
<th>NSS Respondent Satisfaction</th>
<th>Medicine &amp; Dentistry</th>
<th>Education</th>
<th>Global score</th>
</tr>
</thead>
<tbody>
<tr>
<td>7. Feedback on my work has been prompt</td>
<td>2007</td>
<td>39 %</td>
<td>56 %</td>
<td>53 %</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>40 %</td>
<td>64 %</td>
<td>58 %</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>59 %</td>
<td>73 %</td>
<td>67 %</td>
</tr>
<tr>
<td>8. I have received detailed comments on my work</td>
<td>2007</td>
<td>31 %</td>
<td>72 %</td>
<td>59 %</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>33 %</td>
<td>74 %</td>
<td>62 %</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>51 %</td>
<td>81 %</td>
<td>72 %</td>
</tr>
<tr>
<td>9. Feedback on my work has helped me clarify things I did not understand</td>
<td>2007</td>
<td>38 %</td>
<td>60 %</td>
<td>53 %</td>
</tr>
<tr>
<td></td>
<td>2010</td>
<td>41 %</td>
<td>64 %</td>
<td>57 %</td>
</tr>
<tr>
<td></td>
<td>2013</td>
<td>54 %</td>
<td>73 %</td>
<td>67 %</td>
</tr>
</tbody>
</table>
From the information shown in this table it is possible to highlight the following:

i. Medical and dental students show consistently lower satisfaction scores compared to education students and the global score.

ii. Education students display higher satisfaction scores than the global score.

iii. There is a gradual improvement in all three questions for medicine and dentistry, education and the global score, across the 2007 to 2013 period.

iv. Medicine and dentistry made the bigger improvements between 2010 and 2013 while education has done it more gradually from 2007.

v. Despite the improvements, there are still more than 40% of future medical doctors and dentists who have the perception that they are not receiving prompt feedback.

vi. From these medicine and dentistry students, 46% do not see the received feedback as helpful.

As described in the literature, the natural response to the initial low level of feedback satisfaction among higher education students, has been an extra effort to enhance the quality of the feedback information provided by teachers, which appears enormously resource-heavy (Bloxham and Campbell, 2010; Nicol et al., 2014).

Despite the fact that feedback satisfaction in the NSS has increased, there is much work to be done. This because in spite of these better NSS figures, as a result of the described attempts to develop teacher feedback, recent studies show that students’ learning, as a result of the provided feedback, does not improve (Orsmond and Merry, 2011; Evans, 2013; Orsmond et al., 2013).

In this regard, Boud and Molloy (2013 pp. 7-8) have highlighted three conventional assumptions that need to be challenged in order to understand and improve the current feedback provision to enhance students’ learning. These are:

i. Feedback constitutes one-way flow of information from a knowledgeable person to a less knowledgeable person.

ii. The job of feedback is complete with the imparting of performance related information.

iii. A generic model of best-practice feedback can be applied to all learners and all learning situations.
Though these statements are very much interrelated, they will be approached separately for didactic reasons.

First, feedback must ‘not’ constitute a one-way flow of information from a knowledgeable person to a less knowledgeable person. Indeed, scholars now believe that, if feedback is to enhance students learning and develop expertise, it needs to be reconceptualised as a dialogue (Nicol, 2010), a process where students must play an active role (Sadler, 2010), whereby they are given opportunities to construct their own meaning from the received feedback. Thus, students need to engage (Price et al., 2011) with the feedback message, take it, analyse it, ask questions about it, discuss it with others as an ongoing socially-embedded process (Price et al., 2011), and finally, connect it with prior knowledge (Carless et al., 2010; Nicol et al., 2014). The concept of ‘feed forward’ has been introduced to explain this process (Sadler, 1983), on how to improve the students' use of tutors' comments by moving feedback forward. Further, the three phases model to “move feedback forward” later developed by Sadler (1989) can be combined with the Hattie and Timperley (2007) “model of effective feedback” in the following steps: a) According to Sadler (1989), the student needs to start by knowing the standard, goal and criteria for their desired degree of performance or excellence, which in the words of Hattie and Timperley (2007) corresponds to the answers to tutor or peer questions as where am I going? and, what are the goals?, known as ‘feed up’; b) Then, the student must recognise how her or his current performance relates to this standard or criteria, by asking her- or him-self how am I doing?, what progress have I made toward the goal?, recognised as ‘feed back’; c) Finally, the student needs to understand and design a strategy to close the gap between her or his current performance and the goal, by asking where to next?, what activities do I need to undertake to make a better progress?, which corresponds to the notion of ‘feed forward’.

A similar process called the “feedforward interview” has been described by Kluger and Van Dijk (2010) for use in the clinical setting. Surprisingly, none of these studies reference each other.

In essence, by providing ‘feed forward’ feedback, students’ immediate needs of the piece of work or performance are satisfied, delivering high quality information, but at the same time, by encouraging them to engage with the message, they start monitoring their own work to become self-regulated learners (Nicol and Macfarlane-Dick, 2006). This process has been also coined as “sustainable feedback” as it supports students in
self-monitoring their own work independently of the tutor (Carless et al., 2010). Further, as discussed above (Assessment purposes in page 43), fostering students’ metacognition and self-regulation is also central to the concept of assessment as learning (Earl, 2012), where students become active critical thinkers, comfortable with reflection, making sense of the feedback message, relating it to prior knowledge and using it to construct new learning.

However, in the current educational environment where resources to hire new staff are limited and the number of students are increasing (Light et al., 2009), there arises the question of how all these demanding jobs are going to be accomplished (Sadler, 1989; Price et al., 2010). Thus, some scholars have suggested the incorporation of peers to support the feedback process (Boud et al., 1999). This is informed by research showing that peer-feedback’s accessible language is often better understood than that of the tutors’ (Topping, 1998; Boud and Falchikov, 2007). Further, receiving feedback from multiple peers has been reported to be highly reliable (Cho et al., 2006), and more helpful in improving the quality of students’ assignments than receiving feedback from one peer or a single expert (Cho and MacArthur, 2010). Moreover, recent research has shown that formative peer-assessment may be a vehicle to close the gap between the feedback given to students and feedback effectively used by them (Cartney, 2010), and that repeated involvement in dialogic peer-feedback interactions with a self-assessment component would support the process of self-regulation (Carless et al., 2010). On this point, peer-feedback provides a more social learning format, taking learning out of the private domain (Liu and Carless, 2006), which has been seen to achieve higher standards in comparison to the one-way flow practice of tutors telling students about the quality of their work (Sadler, 2010).

Second, the job of feedback is ‘not’ complete with the imparting of performance related information. As mentioned above, students need to take action on the received feedback to learn from it. It can be accepted that the process of feedback might be prompted by a tutor’s words or writing, but the process is not concluded until action by the student occurs (Boud and Molloy, 2013).

Thus, aiming to explore the extent to which undergraduate social work students acted on their assessors’ feedback, Crisp (2007) concluded that providing feedback alone was not sufficient to improve students’ written work. She also noted that assuming that students understood the received feedback because they had not sought subsequent
clarification about comments made, was wrong. In line with this, Higgins et al. (2002) focused on business and humanities students’ understanding of feedback, highlighting that students need to have misconceptions ‘explained’ as soon as possible, and that simply telling them their mistakes, was not enough. Further, a recent Cochrane Collaboration review on the effects of audit and feedback on professional practice and healthcare outcomes, found that when feedback was given both verbally and in writing, and when it includes a clear target and an action plan, it is most effective (Ivers et al., 2012). This reaffirms the conclusion that feedback should not be just ‘giving’ information, and that it crosses the discipline as well as the undergraduate and professional borders.

Casting doubts on whether the practice of feedback as just ‘telling’ is ever to change, Bailey and Garner (2010) interviewed 48 lecturers from across departments and faculties, who stated that despite their university’s commitment to ensure timely and useful formative feedback on assignments, this was not having the intended effect. Lecturers were uncertain about what use students made of the provided feedback, and honestly but worryingly, many seemed to have become “indifferent to the educational value of written feedback” and did it just to comply with institutional policies.

This apathy might be explained by the findings of Evans (2013) in that the demands on the lecturer to support students’ access to and engagement in feedback are ‘huge’. Notably, the lecturers in the former study made clear their doubts of the use students gave to the provided feedback, limiting their role to the provision of feedback, which might explain why students still feel, as expressed in the NSS, that feedback on their work has not helped them clarify things they did not understand. In line with this, Orsmond and Merry (2011) suggested a misalignment in feedback provision as developmental aspects of students’ learning were rarely addressed in tutor feedback. Further in this dissonance, Carles (2006) established that tutors believe that they are providing more detailed feedback and that it is more useful than students do, while Adcroft (2011) found that students perceive they receive feedback much less frequently than tutors perceive they give it.

Ideally then, and following the concept of assessment as learning (Earl, 2012) described above, lecturers should do their best to ‘engage’ students in feedback (Evans, 2013), going beyond the judgement of ‘where’ and ‘what’ to improve (Bloxham, 2009), by helping them to understand the meaning of the feedback by having a dialogue (Nicol,
2010; Orsmond et al., 2013) and encouraging them to take action on it (Sadler, 2010). This should also consider other factors such as tutor’s low credibility among students as a reason for their unwillingness to act on their tutor’s feedback (Orsmond et al., 2005).

However, not all the blame is on the tutors. In the study by Wingate (2010), it became apparent that students paid little attention to their tutors’ feedback comments and did not act upon them. The reasons for this included their low motivation based on the enjoyment of the degree programme they were following, and the self-perception of their ability as writers. The author also highlighted the need, as feedback providers, to pay more attention on the comments given, targeting them differently to high- and low-achieving students. This implies knowing the students better, leading us to the third assumption that one feedback model suits all learners and situations.

Third, a generic model of best-practice feedback can ‘not’ be applied to all learners and all learning situations. The reasons for this are that university-based and workplace environments have variations in contexts, persons and risks, which do not allow for an appropriate application of currently advocated feedback models. Further, the increasing numbers and diversity of university students coming from a wide range of educational experiences are being educated for increasingly diverse practices and workplaces. This does not allow feedback to be based on a common set of assumptions, as was done before (Boud and Molloy, 2013).

A reported quick-fix solution to the above presented NSS low scores for feedback, in which tutors were encouraged to signal and underline their use of anything that seemed like feedback on every occasion they could think of, does not recognise the magnitude of the real problem and the need for change (Boud and Molloy, 2013). These approaches will probably not satisfy students because what they are seeking is a dialogic individual communication (Nicol, 2010; Dowden et al., 2011).

It should also be noticed that quick feedback tricks which are not contextualised may have no effects or even negative consequences, as they should consider the task, the process, the learning setting, the learner’s motivation and self-regulation (Kluger and DeNisi, 1996; Hattie and Timperley, 2007; Kluger and Van Dijk, 2010). For example, Kluger and DeNisi (1996), found in their meta-analysis that feedback effectiveness decreases as attention moves up the hierarchy closer to the self and away from the task motivating the feedback.
It is important to ensure the feedback message is targeted at students at the appropriate level (Hattie and Timperley, 2007). Thus, feedback is not uniform in use and in concept, and when, how, and by whom feedback is delivered matters and the effects are variable as a function of the specificities of the situation (Eva et al., 2012).

1.4.9 Feedback that works

There are many claims that learning and performance can be optimised by the provision of feedback. In the Hattie and Timperley (2007) original review, the effect size of feedback from 12 meta-analyses including 196 studies was $d=0.79$ (twice the average effect of $d=0.40$). In a recently published book, Hattie (2012) upgraded the factors achievements by including now over 900 meta-analyses, where feedback is placed in the top ten influences on achievement, though with considerable variability.

Consistently, another meta-analysis of 131 studies conducted by Kluger and DeNisi (1996), addressed the effect of 36 feedback moderators and found an average effect size of $d=0.38$. The wide range included a maximum effect size of $d=0.55$ for “task feedback about changes from previous trials” as well as in feedback for “not complex tasks”, to a negative effect size of $d=-0.31$ for task feedback designed to “discourage the student”. Similarly, there were ineffective feedback interventions such as ‘praise’ for task performance, and with a weak effect such as physical tasks (negative effect) and following rules tasks. The authors highlighted that most of the variances could not be accounted for by sampling error. Surprisingly, 32% of the feedback interventions had a negative effect size, though only ‘discouraging feedback’ was statistically significant.

In addition, the Cochrane Collaboration review on the effects of audit and feedback on professional practice and healthcare outcomes, referred above, found three studies with large negative effects related to prescription of medicines and laboratory test utilization (Ivers et al., 2012).

Smaller studies on students’ experiences have also helped in identifying feedback perceived as unhelpful to improve learning. Weaver (2006) found that comments which were too general or vague, lacked guidance, focused on the negative, or were unrelated to assessment criteria, were not beneficial.

A systematic review on assessment, feedback and physicians’ clinical performance conducted by Veloski et al. (2006), found that 32/41 (74%) studies that evaluated the independent effect of feedback on physician performance (baseline measurement of
performance, feedback and follow-up assessment), demonstrated a positive impact. The most effective feedback on physician performance was found to be provided by a credible and authoritative source over a number of years. Unfortunately, as the authors stated, the variation in outcome variables precluded any systematic analysis of effect sizes.

From these and other studies, it appears that feedback is more effective when:

- It directs information to enhance self-efficacy in the task and to more effective self-regulation, promoting reflection on actions (Archer, 2010), and stimulating students’ motivation (Wingate, 2010). In this case, feedback is likely to produce remarkable performance improvements (Kluger and DeNisi, 1996). Therefore, students need quality over quantity feedback (Dolmans, 2013) about their performance against clearly defined learning outcomes (Harden and Laidlaw, 2013). This is particularly important for the disadvantaged and low-attaining learners, as they benefit the most from a formative assessment feedback (Black and Wiliam, 1998).

- It is phrased in non-evaluative language as possible. Comments like their performance was ‘totally inadequate’ are not helpful (Harden and Laidlaw, 2013). It should always be remembered that trainees are generally apprehensive about feedback and their fear will only disappear when they realise that feedback is essential for effective learning and the development of competence (Pelgrim and Kramer, 2013).

- It is given in a timely manner (Subramanian et al., 2013), when memory is still fresh for both observer and trainee (Weaver, 2006; Kilminster et al., 2007; Shute, 2008), but allows emotions to be reduced (van der Leeuw and Slootweg, 2013). Harden and Laidlaw (2013) found that providing feedback immediately following an Objective Structured Clinical Examination (OSCE) was a useful and powerful learning experience. Further, prompt feedback has been suggested as a condition for a clinical behaviour to be corrected (Ramani and Krackov, 2012).

- It is provided at the end of the task (terminal feedback) (Walsh et al., 2009), in a private setting (Neher et al., 1992), giving the students some time to reflect on their learning (McMillan, 2011).

- It builds on changes from previous trails containing cues that support learning, drawing attention to discrepancies between the task and the standard (Kluger
and DeNisi, 1996). Thus, Wingate (2010) found that students who had utilised their earlier feedback comments improved in those areas previously criticised.

- It includes an action plan with personal goals and targets that are specific, clear and challenging but overall task complexity is low. Kamp et al. (2013) observed that the impact of peer-feedback at the ‘process’ level according to Hattie and Timperley model (2007) (explained below in this section), can be increased by combining it with individual reflection by means of goal setting with face-to-face discussion.

- It moves away from a trainer-driven monologue to a valuable dialogue (Archer, 2010; Orsmond et al., 2013), where feedback, in a socio-constructivist paradigm (Vygotsky, 1978; Jarvis, 1987), is seen as facilitative, enabling students to make their own revisions and, through dialogue, help students to gain new understandings without dictating what those understandings will be (Evans, 2013). This is hindered by the classic medical hierarchical model (Archer, 2010), in which the cognitivist perspective is associated with a directive telling approach where feedback from an expert provides information to a passive novice (Evans, 2013).

- The feedback provider is a supervisor, peer (Ivers et al., 2012), or a role model near-peer (Nelson et al., 2013) who is aware of the qualities of the discussions that result in successful feedback (Mehta et al., 2013), and that can separate the content from the relationship (van der Leeuw and Slootweg, 2013). When provider and trainee arrange and plan the observation and feedback, a greater benefit can be expected (Pelgrim et al., 2012a). Further, students’ receptivity of the feedback is influenced by the credibility of the provider (Watling et al., 2012a). Accordingly, they will reject feedback from a not credible assessor, or not truly engage in the creation and exchange of informed and accurate feedback (Watling et al., 2008).

- It is provided frequently (Kilminster et al., 2007; Ivers et al., 2012), in stress-free conditions (Rolfe and McPherson, 1995), ensuring all performance evidence is available and first-hand (Harden and Laidlaw, 2013). However, feedback provision in busy, time-constrained clinical settings is frequently overlooked (Archer, 2010; Sabey and Harris, 2011).

- It is given both verbally and in writing (Ivers et al., 2012) in a comprehensible language to the student (Higgins et al., 2002), and in a way that facilitates the development of self-assessment (reflection) in learning (Nicol and Macfarlane-
• There are perceived low rather than high levels of threat to self-esteem, because low-threat conditions allow attention to be paid to the feedback. It should also encourage positive motivational beliefs (Nicol and Macfarlane-Dick, 2006).

• Combined with other interventions such as effective instruction in classrooms (Kluger and DeNisi, 1996), reminders (Ivers et al., 2012), educational programmes and practice guidelines (Veloski et al., 2006).

In order to analyse these factors in more detail, the Hattie model of feedback will be used as a framework to identify conditions to maximise the positive effects of feedback on learning (Hattie and Timperley, 2007; Hattie, 2012). Before going deeper in the model, it should be stated that a broader aim is to create, as called by Sargeant et al. (2011) a “culture of improvement” where sharing and seeking feedback to improve performance is the norm of work and learning.

Thus, the rationale of Hattie’s feedback model is to reduce discrepancies between current understanding and performance and a given goal. However, from the information presented above, we already know that not all strategies to reduce this gap and enhance learning are equally effective. Accordingly, the first notions to start with in Hattie’s model of feedback are the answers to three major questions asked by a tutor or peer, as previously presented. By recapitulating, these are:

a) Where am I going? (‘feed up’ stage)
b) How am I going? (‘feed back’ stage)
c) Where to next? (‘feed forward’ stage)

The answers to these questions can be very effective to reduce the gap, but can also be otherwise. This depends on the ‘level’ at which each of the three questions is focused. According to Hattie and Timperley (2007), these levels are:

1) ‘Task and product’
2) ‘Process’
3) ‘Self-regulation’
4) ‘Self’

These four levels are discussed below and where appropriate, they are complemented with evidence from other studies.
Feedback about the ‘task and product’ level is also known as ‘corrective feedback’ or ‘knowledge of results’. It can be very powerful in novice learners, and when it is more information focused (e.g. correct or incorrect responses), leads to the students to get and provide more or different information relevant to the task, and building more task knowledge. It corresponds to the answer(s) how well has the task been performed? Is it correct or incorrect? It is not only critical, but is the foundation to build effective ‘process’ and ‘self-regulation’ levels (Hattie, 2012). However, the learner needs an explanation as to what she or he did or did not do to meet the expectations. Simply marking an examination (Rolfe and McPherson, 1995; Shute, 2008) or telling the students what went right or wrong, without explanation is less likely to improve their performance (Harden and Laidlaw, 2013).

For this to occur in the clinical setting, the ‘task’ performance must be observed. This makes a big difference, as if the feedback is to be given about a written ‘task’, the tutor or peer feedback provider does not need to observe the procedure, but its result. Unfortunately, performance observation in health education is low (Williams et al., 2003; Kogan and Hauer, 2006; Norcini, 2011), and, as discussed below (Clinical workplace-based assessment and the role of peers in page 81), high proportions of students frequently report difficulties in finding a clinical assessor (Quantrill and Tun, 2012).

The second level focuses the tutor or peer feedback to the ‘process’ used to complete the task. It looks for the student to understand and connect ideas, providing strategies to identify mistakes and transfer the learning experience from these mistakes to other more difficult tasks. This move from ‘error detection’ to ‘error correction’ is highly desirable and has been related to student motivations to close the gap between their current performance and the goal (Boud and Falchikov, 2007). The ‘process’ feedback that contains specific narratives is then more effective to enhance deep learning and gets higher satisfaction scores from recipients (Overeem et al., 2010), as it is more readily integrated and assimilated (Sargeant et al., 2008). The ‘error correction’ with frequent, specific and precise narratives, is particularly important for those novice students who cannot judge their performance, or competency, accurately (Hauser and Bowen, 2009), as they cannot learn by simply observing their tutors (Hendricson and Kleffner, 1998).

Whilst Grieveson et al. (2011b) reported the high perception of dental trainees in that feedback received was provided in a supportive way, had a positive effect on their
training, encouraging them to be reflective in their clinical practise, not all feedback practices reach this level. A study by Mutch (2003) in a university business environment, concluded that feedback with no indication of how deficiencies could be addressed was of little value as it could only be used by those students who recognise what the implied developments were. In line with this, a big study including 460 staff and 1740 students in the eight publicly funded universities in Hong Kong (Carless, 2006), found that 38.4% of the tutors thought students were often given detailed feedback which helped them improve their next assignment, whilst only 10.6% of students responded in the same way. Further, 37.8% of students felt that feedback was rarely followed by actions to improve student learning, as opposed to 16.1% of tutors.

The third feedback level focuses on the student ‘self-regulation’, or, in other words, to the student’s monitoring their own learning process towards the goal. The aim here is to develop the student’s capability of internal feedback or self-assessment, as when they do, they reduce discrepancies between where they are in their learning and the desired outcomes, more effectively. For example, feedback at this level can be focused to help the student to self-evaluate by asking reflective or probing questions on ‘when’, ‘where’ or ‘why’ related to the task. Students who learn the metacognitive skills of self-assessment, develop their self-appraisal (facility to review and evaluate their competences through reflection on action) (Archer, 2010) and self-management (monitoring and regulating their behaviour through planning, error correction, and fixing strategies) skills as well, and by doing so, they know how and when to seek and receive feedback from others. As discussed above (Reported benefits of peer-assessment in page 54), many educational initiatives include in their programme the combination of a self- and a peer-assessment component (Falchikov, 2007) aimed at helping students to take more responsibility for their own learning (Black and Wiliam, 1998). Further, McMillan (2011) suggests asking the student for a self-appraisal of their performance – identify aspects that went well, areas of difficulty and possibilities for change - before offering feedback in order to develop their self-reflective skills.

The fourth level is feedback directed to the ‘self’ as a person. It is important to be aware that personal feedback, commonly considered as ‘praise’, such as ‘you are a great student’ or ‘great effort’, often directs the attention away from the task, process, or self-regulation. Precisely because of this, praise and feedback about the learning should be kept separate. Praise includes little information about the task performance and therefore does not help in answering the three feedback questions. Thus, whilst Skipper
and Douglas (2012) found that ‘self’ feedback was likely to have a negative effect particularly in those students who are not successful, Boehler et al. (2006) reported that students who received general compliments (praise) provided higher satisfaction rating compared to those who received feedback. However, only the latter group increased their performance. Further, Parkes et al. (2013) tested the effectiveness of the frequently used (Kogan et al., 2012) “Sandwich Feedback Technique” – Praise; Critique; Praise - (Davies and Jacobs, 1985) in a peer-feedback exercise of 3rd year medical students, and found that students believed feedback sandwiches positively impacted their subsequent performance, but in reality they did not. Feedback should not undermine self-esteem, but should not simply consist of praise (Kluger and DeNisi, 1996).

An interesting point raised by Bing-You and Trowbridge (2009) is that the current students’ dissatisfaction with feedback (referred to medical students in the USA), “may reflect a greater desire for praise than for constructive information to help them learn”. However, moving away from ‘perceptions’ and focusing on students’ learning, feedback needs to move from the task (what do I know and what can I do?) towards the process or understandings necessary to learn the task (what do I not know and what can I not do?), and from here to self-regulation about continuing beyond the task to more challenging goals (what can I teach others and myself about what I know and can do?) (Hattie, 2012). However, in the medical education literature there still are calls to combine both compliments and feedback (Boehler et al., 2006).

Figure 1-6 Graphical interpretation of the peer-feedback characteristics that could have a positive or negative influence. This could be used for peer-assessment training purposes.
Some authors have stated that a regular self-, peer- and/or multisource assessment-feedback strategy (Aronson, 2011), with students actively working in different roles as part of a team (McMillan, 2011), facilitates and is crucial to deeper learning (Boursicot et al., 2011; Manogue et al., 2011; O’Donnell et al., 2011). In line with this, Sweet et al. (2008) perceived that students who worked together and engaged in discussions about what they do in the clinic, can be expected to develop a deeper learning with changes in their perspectives.

### 1.4.10 Reflection and learning

Reflection, reflective learning, reflective writing and reflective practice are being increasingly used not only in higher education but also in professional development (Mann et al., 2009; Gonzalez et al., 2013). However, there are some differences in the views of how reflection is defined (Sumson and Fleet, 1996), though consistent with the Latin origin of the word as “to bend” or “to turn back” (Moon, 2007). A few selected definitions of reflection are:

John Dewey (1938), philosopher, psychologist, and educational reformer.

“An active, persistent and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and further conclusions to which it leads...it includes a conscious and voluntary effort to establish belief upon a firm basis of evidence and rationality’ and the further conclusion to which it tends”.

David Boud et al. (1985 p. 19), researcher and teacher of adult higher and professional education.

“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. It may take place in isolation or in association with others”.

John Sandars (2009), medical doctor, lecturer in medical education.

“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”.
Jennifer Moon (Moon, 2013), educational developer and professional trainer.

“A form of mental processing - like a form of thinking - that we use to fulfil a purpose or to achieve some anticipated outcome. It is applied to relatively complicated or unstructured ideas for which there is not an obvious solution and is largely based on the further processing of knowledge and understanding and possibly emotions that we already possess”.

From these definitions, it can be noticed that Boud et al. (1985) and Sandars (2009) are focussed on the process of reflection, while the latter adds the temporal component stating it can occur “before, during and after”, as well as the prospective use of the outcome. Similarly, Moon (2013) mentions the purpose of reflection to anticipate an outcome, and as Boud et al. (1985) do, she considers emotions. Finally, Dewey (1938) presents a holistic view of the process of reflection for learning (Eraut, 1995), highlighting its voluntary character.

On a further analysis of these definitions, it can be notice that they do not make an explicit difference between the previously described Schön (1983) concepts of reflecting during the event (reflection-in-action) or after the event (reflection-on-action) (described in Reflective practice learning in page 35). Probably the definition by Sandars (2009) is the only one clearly focusing on reflection-on-action, when he mentions ‘turn back’ thoughts. Though debatable, this is in agreement with other medical researchers who have stated that in “medical education”, most reflection is on-action (Aronson, 2011), if at all (Branch and Paranjape, 2002).

This might be what in reality happens during medical education. That is, exclusively adopting the previously discussed Kolb’s model (1984) where students only reflect ‘after’ their experiences. As Finger and Asun (2001 p. 45) put it, they are reflecting following a ‘mistake’, that is, “learn by trial and error”. However, this approach would hinder students’ clinical reasoning since this is conceptualized as reflection-in-action (Schön, 1987). Further, as reflection-in-action involves reflecting and doing it in the midst of action (Stegeman et al., 2013), “it implies that the professional has reached a stage of competence where she or he is able to think consciously about what is taking place and modify actions virtually instantaneously” (Hatton and Smith, 1995). Supporting this approach, Mamede et al. (2007), found that complex and unusual clinical cases led medical residents to switch from automatic (non-analytical reasoning)
to reflective reasoning.

It should be noted that this is not a call to abandon Kolb’s learning cycle; on the contrary, this is a request for a cultural change to include in the teaching of our students’ a requirement for them to reflect ‘while’ practicing and ‘after’ practicing. As presented by Boyd (2008), dental practice needs both reflection-in-action and reflection-on-action, to prepare students for a lifetime of professional development (Frankford et al., 2000). This entails the creation of new ways of thinking and acting about practice problems, where the feedback process has been mentioned as key to successful growth and learning (Lee and Caffarella, 1994). In the same line, other researchers have called for encouragement of reflection-in-action in health professional education in order to achieve truly effective feedback (Archer, 2010).

The need for reflection to accomplish effective learning was emphasised by Dewey some time ago (Dewey, 1909; Dewey, 1938), not forgetting the ancient discussions of Plato (Plato, 1952 pp. 684-685 [720]) regarding the “thoughtful diagnosis and reasoning about treatment and explanation to the patient” (Ericsson, 2008) by a freeman doctor. As well as the previously described models of experiential learning (Kolb, 1984) and reflective practice (Schön, 1983) (pages 33 and 35, respectively), Mezirow’s (1991) constructivist model of transformative learning, also encourages a reflective interpretation of experiences to “make meaning” of them, as fundamental to what learning is all about.

Despite the known benefits (Branch and Paranjape, 2002; Mann et al., 2009) and the requirements to incorporate reflective teaching (American Commission on Dental Accreditation CODA, 2010; Australian Dental Council, 2010; General Dental Council, 2012a), its implementation is challenging (Boud and Walker, 1998; Grant et al., 2006; Mann et al., 2009; Sargeant et al., 2009), in dental traditional curricula (Pee et al., 2002; Woodman et al., 2002; Ashley et al., 2006). Further, the reality seems to be that while feedback to promote reflection is not used often enough, reflection is probably used even less (Branch and Paranjape, 2002), despite students’ recognition of the need for reflection especially after starting clinical practice (Carr and Johnson, 2013).

In the field of healthcare education and practice, reflection on one’s own experiences is generally recognised as a core, critical competence (Mann et al., 2009; Lindström et al., 2011) and the foundation of purposeful learning (Amulya, 2004). However, as Jarvis has pointed out (Jarvis, 1987 p. 84), some individuals do not learn from experiences.
Precisely on this, a key finding in Tsang and Walsh (2010) study on last year oral health students, was that prior to implementing a clinical reflective learning programme, “clinics were attended and then forgotten”. It is today accepted that knowledge and practical skills alone are not enough to become a medical professional (Boenink et al., 2004), as the critical analysis of lived experiences, in order to understand their broader context, allows the learner to actively integrate the new resulted knowledge (Moon, 2013). By developing the habit of exploring and being curious about our own struggles, uncertainties, and breakthroughs, we open the possibility of a purposeful learning, not from external sources, but from our own work and lives (Amulya, 2004).

A number of studies explain the effects of reflection on the learning processes of students and trainees. For example, reflection has been reported to improve, develop or facilitate a range of attributes, including critical thinking (Phan, 2008; Mitchell et al., 2009) clinical reasoning skills (Baernstein and Fryer-Edwards, 2003; Sandars, 2009), diagnostic reasoning abilities (Sobral, 2000), particularly in complex and unusual cases (Mann et al., 2009), medical-humanism skills (Epstein, 1999; Wiecha et al., 2002; Gracey et al., 2005), as well as technical skills, evidence-based decisions (Epstein, 1999) and professionalism (Mofidi et al., 2003; Stern and Papadakis, 2006; Phan, 2008).

Pedagogically and from a constructivist viewpoint (Knowles et al., 2012), reflection allows the integration of new learning into existing knowledge and skills (Mann et al., 2009), promoting self-regulated learning activities (Boud et al., 1985; Grant et al., 2006) which assist and encourage professional development (Tsang and Walsh, 2010; Lindström et al., 2011). This is accomplished by increasing the self-awareness of students and trainees in the learning processes (van den Boom et al., 2007; Sandars, 2009), and so enhances the learning opportunity (Sobral, 2000; Lonka et al., 2001; Cleary and Sandars, 2011) through a deeper approach to learning (Sobral, 2005; Sandars, 2009; Tsang and Walsh, 2010).

By way of contrast, there are some studies that report no measurable academic improvement after reflective writings (Lew and Schmidt, 2011), or problems with, or adverse effects from, forced reflection. Thus, an “instrumental or rule-following approach to reflective activities” has been suggested as leading to reflection without learning and inappropriate levels of disclosure with strategic responses, tension between public and private reflections, hostility and even moral questions (Boud and Walker,
1998; Boud, 1999; Hobbs, 2007; Vivekananda-Schmidt et al., 2011), as well as the necessity (Jonas-Dwyer et al., 2013), with its associated anxiety to find time for reflection (Burnard, 1995; Dornan et al., 2002; Pearson and Heywood, 2004).

Reflective learning does not happen intuitively, spontaneously, voluntarily, or even devoutly (Grant et al., 2006; van den Boom et al., 2007; McDonald, 2010; Chambers et al., 2011), especially in the case of new generations of students (Sandars and Homer, 2008; Wald et al., 2012). Some students believe that learning the technicalities of practice are of the greatest, or only importance, and they do not attempt to reflect or learn from experience (Powell, 1989). Despite this, reflection is an ability that can be taught (Boenink et al., 2004; Wald et al., 2009) yet it requires training, development and regular practice (Driessen et al., 2005; Nicol and Macfarlane-Dick, 2006; Sandars, 2009; Oosterbaan et al., 2010; Aronson, 2011; Aronson et al., 2012).

One approach that has shown to support the development of individual reflective skills is regular peer-evaluation and face-to-face peer-feedback (Pee et al., 2002; Wallman et al., 2008; Sandars, 2009; McDonald, 2010; Subramanian et al., 2013), with clear goal settings (Kamp et al., 2013). If the process of reflection through peer-assessment and feedback can help to reframe and organise theory, then reflection becomes a vehicle for effective learning (Kaufman and Mann, 2010).

1.4.11 Clinical workplace-based assessment and the role of peers

The assessment of clinical competence is a critical issue as it relates directly to the quality of patient care (Nulty et al., 2010; Albino et al., 2012). The introduction and later development of the Objective Structured Clinical Examination (OSCE) (Harden and Gleeson, 1979; Schoonheim-Klein et al., 2008; Schoonheim-Klein et al., 2009), to assess students at the “Shows how” dimension of Miller’s model (1990) (graphically represented in Figure 1.5 in page 46), after a near exclusive reliance on using written and oral examinations with subjective standard settings and poor reliability (Norcini and McKinley, 2007), contributed to a significant improvement in the reliability of authentic competence tasks assessment (Norcini, 2005; van der Vleuten and Schuwirth, 2005; Norcini et al., 2011).

However, what students “show how” to do in controlled high stakes examinations like the OSCE (showing competence), does not assure they will perform competently in actual practice (Rethans et al., 2002). As a result, a challenging step has now been taken
to move the assessment venue from the relatively homogeneous setting of structured examinations, back to the uncontrolled and heterogeneous real world, assembling “teaching, learning, and assessment” in one single place, that is, the workplace (Norcini and Burch, 2007).

Though more research is needed (Kogan et al., 2009), this process has been encouraged by studies showing good validity and reliability for some newly developed, less standardised methods of direct observation workplace-based assessment (WPBA) that focus on the “Does” top end of Miller’s pyramid (Miller, 1990), while trainees are in the real clinical setting of patient care (Norcini and McKinley, 2007).

Thus, these new assessment methods are the outcome of a thoughtful attempt during the last two decades to design structured and standardised assessment forms (questionnaires) of the quality of observed students’ behaviours (Cantillon and Wood, 2010) but in a naturalistic setting (Boursicot et al., 2011), with the advantage of allowing for immediate formative feedback (Norcini and Burch, 2007). The most commonly found WPBA tools in the literature, and therefore subsequently described, are the Mini-Clinical Evaluation Exercise (mini-CEX), Direct Observations of Procedural Skill (DOPS), Longitudinal Evaluation of Performance (LEP), Case-base Discussion (CbD), Multi-Source Feedback (MSF) which includes the Patient Assessment Questionnaire (PAQ) and the Mini-Peer Assessment Tool (mini-PAT).

**Mini-Clinical Evaluation Exercise (mini-CEX)**

The mini-CEX, designed by the American Board of Internal Medicine, is used to assess trainees while conducting a clinical consultation (Norcini et al., 1995), generally of an inpatient, outpatient or emergency department settings (Norcini et al., 2003). Thus, the tutor observes the learner while interviewing a patient and/or conducting a physical examination, and judges this and her or his professionalism, clinical judgment, counselling, communication, organisation and efficiency, as well as the overall clinical competence. The trainee then summarises the encounter to the tutor by providing the diagnosis and/or treatment plan (Cantillon and Wood, 2010).

The encounter is recorded and scored on the mini-CEX form (Appendix 5) by the trainer, who uses this information to provide the trainee with structured feedback during debriefing (Cantillon and Wood, 2010). The encounter takes about 10 or 15 minutes with 5 minutes for feedback. Typically, the trainee would be assessed on several
occasions by different trainers, so that bias from a single assessor is reduced. Further, multiple mini-CEX on different patients addresses the case specificity of clinical performance (Norcini, 2011).

The mini-CEX is intended to identify the few trainees whose performance is totally unsatisfactory and to provide the remainder with the opportunity for ongoing formative assessment and feedback (Norcini and Burch, 2007). It has also been designed to ensure that trainees’ clinical skills have been observed and assessed by staff members (Norcini, 2011).

There are copious studies on the use and characteristics of the Mini-CEX mainly in medicine (Norcini et al., 1995; Norcini et al., 1997; Norcini et al., 2003; Hill et al., 2009; Cook et al., 2010; Dewi and Achmad, 2010; Pelgrim et al., 2012b; Al Ansari et al., 2013; Alves de Lima et al., 2013; Liao et al., 2013; Weston and Smith, 2014), but a few in dentistry as well (Millett, 2011; Kalsi et al., 2013).

**Direct Observations of Procedural Skill (DOPS)**

The DOPS is a variation of the mini-CEX, designed by the Royal College of Physicians to assess and provide feedback on practical clinical procedures (Wragg et al., 2003). Just as with the mini-CEX, trainees are observed and assessed by a staff member while practicing a clinical procedure on real patients. The scoring and the provision of feedback are similar to that of the mini-CEX. However, DOPS encounters are usually longer than those of the mini-CEX as they require the duration of the procedure being performed (Wilkinson et al., 2008). Thus, DOPS assessment will normally review the indications for the procedure being performed, how consent was obtained, whether appropriate analgesia was used (if needed), technical ability to perform the procedure, professionalism, asepsis, awareness and management of complications, as well as the overall competence (Appendix 6) (Cantillon and Wood, 2010). Although the original DOPS format was designed as a generic tool that could be used in any clinical procedure (Wragg et al., 2003), there are today large number of DOPS formats to assess different specialties both in medicine and dentistry (The Royal College of Surgeons of England, 2014).

Though not as many as with the mini-CEX, there are studies reporting the use of DOPS in medicine (Wilkinson et al., 2008; Davies et al., 2009; Ahmed et al., 2011; Mitchell et al., 2011; Bindal et al., 2013; Cobb et al., 2013; Delfino et al., 2013; Watson et al.,
2014). However, and although Kalsi et al. (2013) included DOPS in their review of several WPBA methods for foundation and postgraduate dental training, and that the UK Intercollegiate Surgical Curriculum Programme (ISCP) presents twelve different specialty DOPS (The Royal College of Surgeons of England, 2014), there are no published studies in dentistry reporting its use. It can be speculated that this is probably due to the series of published reports directed by Linda Prescott-Clements on competency-based assessment of postgraduate dental training in Scotland (Prescott et al., 2002). Thus, she developed the Longitudinal Evaluation of Performance (LEP) and published it before the appearance of DOPS.

**Longitudinal Evaluation of Performance (LEP)**

As mentioned above, Prescott-Clements et al. (2002) introduced a method of clinical assessment, the Longitudinal Evaluation of Performance, which considered, as identified by the authors, three major challenges that clinical educators face when designing competency-based systems of assessment. These were i) the need to assess different areas of clinical competence; ii) the necessity of alignment between assessment and training objectives, and iii) the type and focus of the assessment method.

Consequently, as it can be appreciated from Appendix 7, the LEP eight domains look like a combination of the mini-CEX and DOPS, as, besides generic skills like “professionalism” and “communication”, it includes for example “consultation skills” and “clinical judgement and diagnosis” typically from the mini-CEX, but also “technical ability and manual dexterity” more distinctive of DOPS. A latter version of the same instrument replaced the original last question “Overall competence” for the “Trainee’s insight into performance” in order for trainees to develop this competence with experience and regular feedback. The satisfactory implementation of this improved version of the LEP tool in postgraduate dental trainees in Scotland has been recently published by Prescott-Clements et al. (2011). Thus, the LEP has been reported to combine a strong formative approach through continuous assessment and systematic feedback, which, within a wider assessment programme, contributes towards a summative decision of competence (Prescott-Clements et al., 2008).

A slightly modified version of the LEP, named “Dental Evaluation of Performance” (D-EP), has been reported to be part of the workplace based assessment tools for the dental foundation training in the English Mersey deanery (Grieveson et al., 2011a), and more
recently in other 12 deaneries in England and Northern Ireland (Kirton et al., 2013).

**Case-base Discussion (CbD)**

The CbD is the UK variation of the Chart-Stimulated Recall (CSR) developed in the United States to be used in the context of emergency medicine (Maatsch et al., 1983). In both cases, the trainees chose several patient records whom they have recently seen and in which they have made notes (Cantillon and Wood, 2010). Subsequently, in an encounter of approximately 20 minutes (Norcini and Burch, 2007), the assessor selects one record for discussion and asks the trainee to describe the case and asks clarifying questions directed to know why the trainee acted as she or he did (Norcini, 2011). Upon relevant details of the case emerge, the assessor focuses the discussion on the trainees’ diagnostic reasoning, investigation, the rationale for choosing a certain action and their awareness of differential diagnoses, ethical issues, follow-up and future planning. Further, since patient records are available at the time of the encounter, record keeping can also be assessed as the CbD forms usually contain a domain for this purpose (Appendix 8) (Norcini and Burch, 2007).

It is expected that trainees participate in multiple encounters so that several clinical cases are discussed with different trainers (Setna et al., 2010) in order to reduce any bias caused by case specificity (Cantillon and Wood, 2010).

Extensive studies have reported the use of CbD in foundation medical doctors (Mitchell et al., 2011; Mitchell et al., 2013), as well as in medical specialty trainees (Setna et al., 2010; Sabey and Harris, 2011; Mehta et al., 2013). Further, the use of CbD has also been reported in dental foundation training (Grieveson et al., 2011a; Kirton et al., 2013).

**Multi-Source Feedback (MSF)**

The idea behind MSF is to directly assess “routine practice” which has been described to be much harder to assess compared to single encounter assessment (Cantillon and Wood, 2010), such as the ones previously described. Thus, MSF, also known as 360-degree assessment, represent the systematic collection of colleagues’ and patients’ perspectives in order to assess performance and to provide feedback to students and trainees. In the case of patients, they are usually asked to rate the trainee’s communication skills and professionalism (Goldie, 2013) using specifically developed forms (Appendix 9) (Hurst et al., 2004).
A number of different tools regarding peer-assessment have been reported. Evans et al. (2007) used the Objective Structured Assessment of Technical Skills (OSATS) scale, developed to assess technical skill of surgical trainees, to compare single occasion peer-assessment scores of 38 postgraduates trainees, specialising in oral and maxillofacial surgery, to those assigned by trainers, following a third molar surgery.

Lurie et al. (2006) presented an online version of the Peer-Assessment Protocol (PAP), originally developed by Dannefer et al. (2005), to assess behavioural professional competencies like “work habits” and “interpersonal sensitivity” of undergraduate medical students. Recently, Mackillop et al. (2011a) reported the feasibility of generic on-line Multi-source feedback questionnaire for revalidation of UK career-grade doctors.

Archer et al. (2005; 2010) presented the 24-item Sheffield Peer Review Assessment Tool (SPRAT), which reliably, yet feasibly allowed senior medical paediatricians in training to assess each other’s clinical competencies. The same group (Archer et al., 2008) reported later the validation of a simplified SPRAT instrument, namely the 16-item Mini-Peer Assessment Tool (mini-PAT) (Appendix 10).

In the mini-PAT, peers’ assessment is given anonymously (Norcini and Burch, 2007). Typically the trainee selects eight assessors representing a mix of senior supervisors, trainee and nursing colleagues, clinic staff and so on, who are all requested to complete the mini-PAT and to return them to a central location for processing. The trainee also self-assesses, completing the same form. Subsequently, all questionnaires data are amalgamated and presented to the trainee by a supervisor in a manner that she or he can compare the self-assessment with the mean ratings of the peer assessors. Trainee and tutor then agree on strengths and what aspects of clinical, professional or team performance need more work. It is normal practice in the UK Medical Foundation Programme to perform this assessment twice a year during the duration of the training programme (Norcini, 2011).

The mini-PAT has been used in the medical foundation assessment programme in conjunction with some of the previously described WPBA forms (Archer et al., 2008; Davies et al., 2009; Mitchell et al., 2011; Mitchell et al., 2013), in medical surgical trainees (Pereira and Dean, 2009; Eardley et al., 2013), and in postgraduate pharmacists (Davies et al., 2013). In dental education, the use of the mini-PAT has been suggested (Mattheos et al., 2009) but there are no publications reporting its use. However, as part
of a multisource assessment the Patient Assessment Questionnaire (PAQ) has been used for Foundation Dental Training in the UK (Grieveson et al., 2011a; Kirton et al., 2013).

Studies reviewing these and other WPBA methods have shown positive and sufficiently reliable and valid results to be used for formative purposes (van der Vleuten and Schuwirth, 2005; Norcini and Burch, 2007; Cohen et al., 2009; Davies et al., 2009; Kogan et al., 2009; Boursicot et al., 2011; Norcini et al., 2011; Pelgrim et al., 2011; Driessen et al., 2012; Donnon et al., 2014).

However, a number of potential concerns over the current practice of WPBA have also emerged (Spencer, 2003; Pereira and Dean, 2009; Saedon et al., 2010; Archer and McAvoy, 2011; Beard, 2011; Al-Kadri et al., 2013; Driessen and Scheele, 2013; Ferguson et al., 2014). A review paper by Miller and Archer (2010), pointed out that very few published articles explore the impact of WPBA on trainees’ education and performance. They showed that MSF can lead to performance improvement while there was no evidence that DOPS, Mini-CEX or CbD did so. In a later study of 52 general medical practitioner specialist trainees conducted by Sabey and Harris (2011), 74% reported Multi-Source Feedback to be the most useful tool, whilst 35% did for CbD, 14% for DOPS and only 10% for Mini-CEX. While 55% found WPBA useful as a learning tool, only 45% reported the use of WPBA for summative assessment to be helpful in identifying a struggling doctor. Trainees place a low value on rating scale scores and they perceive a lack of honesty in assessments, as well as bias and a “box ticking” attitude that undermines the credibility of WPBA. In another recent study (Quantrill and Tun, 2012), 95% of trainees reported problems in finding an assessor to complete their WPBA. Although feedback should be at the very heart of WPBA, only 43.9% of trainees received immediate feedback and 22% received it one week later. The reason for this lack of feedback has been attributed to senior trainees or clinical supervisors with limited time or incentive to give due attention to WPBA (Quantrill and Tun, 2012). Thus, Saedon et al. (2010) emphasised the need to train staff on the great importance of quality feedback (including theory), and trainees around the need for feedback, how to give and receive it. Similarly, Bindal et al. (2011) have called for additional training especially on the value of WPBA and consultants to have protected time in their job plans for training.

Additionally, Norcini and Burch (2007) reported the apparent poor Faculty participation...
in formative assessment and feedback strategies as probably the most significant limiting factor currently identified in WPBA. One challenge set by the authors is to make current WPBA methods more user-friendly by their modification and simplification, especially as they are designed to be used in busy clinical settings.

Fortunately, published studies using WPBA in dentistry have shown somehow better outcomes. Thus, Prescott-Clements et al. (2008) reported positive results to support the validity, educational impact and feasibility of the LEP, using almost 10,000 completed forms submitted over 2 years by 201 postgraduate trainees. Later, using the subsequent version of the LEP, previously described, the same researchers (Prescott-Clements et al., 2011) once again obtained encouraging results when tutors assessed and provided feedback to 139 postgraduate dental trainees to develop their insight into their performance.

In another study of 41 Foundation dental practitioners, Grieveson et al. (2011a) surveyed their perceptions of the effectiveness of three WPBA methods used during their training: the Dental Evaluation of Performance (D-EP), Case-based Discussion (CbD) and Patient Assessment Questionnaire (PAQ). Overall 84.1% of trainees felt that WPBA helped them improve patient care (96% with CbD). More specifically, 81.6 of trainees felt that feedback provided after being assessed by tutors with DEP highlighted areas to develop their learning better than when CbD (76.8%) or PAQ (60.9%) were used. Further, a large majority (>79%) found that feedback gave them an insight into their own development needs.

A similar but larger study of 359 Dental Foundation trainees published later by the same group (Kirton et al., 2013), showed that the overall experience of WPBA was positive and played an important role in trainees’ learning and building confidence. Interestingly, in a further question not included in the initial report, the majority of trainees (>62%) felt that feedback received from tutors enabled them to be reflective in their clinical practice. Further, trainers and trainees found CbD the most beneficial of all the WPBA used tools. However, in agreement with reports from medical education (Saedon et al., 2010), the authors highlighted the need for comprehensive training in the WPBA tools to ensure their efficacy.

Despite the widespread dissemination of WPBA practices in medical as well as in dental foundation programme and postgraduate training, there is a lack of literature of its use in undergraduate medical and dental education. Among the very few studies
available, Nesbitt et al. (2013) surveyed 288 year-4 medical students in two consecutive cohorts, to collect their experiences on the use of WPBA. A total of 46% (133) of the students found that feedback from the WPBA was useful (19% disagree or strongly disagree). An interesting question assessed students’ perception on whether WPBS were a useful way of making sure that supervising doctors spent time with them while with patients and discussing cases: 42% agree or strongly agree, while 36% disagree or strongly disagree.

More in the psychometrics, Hill et al. (2009) estimated the validity and reliability of a modified mini-CEX when staff assessed medical undergraduate students. The authors found good overall utility of the modified mini-CEX for assessing aspects of the clinical encounter. Strengths included fidelity, wide sampling, perceived validity, and formative observation and feedback. Reliability assessed by means of Cronbach alpha was $\alpha = 0.73$ for 15 encounters.

While the mini-CEX has been described as a single WPBA encounter tool (Cantillon and Wood, 2010), Playford et al. (2013) used it as a staff performed longitudinal assessment of medical undergraduate students over a three-year period. They analysed 5,686 mini-CEX completed forms and found a Cronbach alpha reliability of $\alpha = 0.80$. Marks were significantly affected by the grade of the marking tutor, difficulty of the clinical encounter, and the clinical discipline. An increase in mini-CEX marks over the course of the academic year was also noticed, especially during the initial formative months. The authors concluded that the longitudinal assessment identified and contributed to students’ skills development. Further, students actively participated in their own development by using the mini-CEX in excess of course requirements and seeking harder markers.

Finally and closer to the aim of the present study, Bennett et al. (2012) piloted the utility and students’ acceptability of the mini-CEX as a framework for peer-feedback. Thus, 40 undergraduate medical students, working in pairs (alternating learner and assessor roles), undertook two peer mini-CEX evaluations whilst on clinical attachments. Accordingly, they marked and identified areas for improvement in their peer’s performance. The results showed that students liked to compare their performance, being at the side of the examiner, and receive useful peer-feedback. However, they questioned the validity of the peer-feedback, and asked for more guidance regarding the standards for peer-assessment. The authors reported the mini-CEX to be a useful
framework for formative peer-feedback of undergraduate students. Further, peers were able to identify those areas of good performance as well as those where improvement was required. However, in contrast to the provision of useful feedback, peers did not score less than 4 out of 5, which, according to the authors, was unlikely to reflect the real range of the students.

1.5 Literature review conclusions

The literature on education is complex, diverse, rich in content and far more extensive than can all be included in these pages. At the same time, many opinions, principles and theories are argued and contra-argued constantly which enriches it even more. This has allowed the accumulation of a huge body of educational theory in the social sciences that could be used to inform daily theoretical as well as clinical teaching practice. However, the present review seems to show that the relationship between this rich accrued knowledge and healthcare education could be improved for the benefit of our teaching. It appears that healthcare academics, who share and contribute to this educational knowledge, struggle to demonstrate to fellow practitioners the usefulness of the guiding principles it can provide to inform their teaching practice (Kneebone, 2002). Further, educational textbooks are, arguably, often associated with laboratory experiments and thus, far from the reality of classroom practice (Harden and Laidlaw, 2013). Perhaps debatably, medical education feels privileged and sets itself apart from the monotony of higher education (Swanwick, 2011), as if it had nothing to learn from other fields of education.

The strong evidence supporting the benefits of feedback with very specific characteristics, as well as the implementation of peer-assessment practices focused on its learning potential, not just as a marking exercise, that are largely available in the educational literature, would lead one to expect this to be reflected in the dental education literature. As shown above, this is not the case. However, this does not necessarily mean these practices are not being implemented in dental education, but if we only take the example of the NSS on dental students’ feedback perception compared to that from education students (Table 1-2 in page 64), it appears as if the evidence is not being used to its full potential.

The wide theoretical and practical background of peer-assessment from other areas of education, as presented above, provides support and underpins the further study of peer-
assessment and peer-feedback in dental students as a possible learning experience. The high contact time dental clinical partners share in their respective workplace, while practicing their skills or during patient treatment makes it an ideal environment for them to help each other through formative peer-assessment and peer-feedback, provided there has been careful and rigorous planning, including student training.

The same literature highlights the challenges dental students’ peer-assessment experience might encounter, among these are reliability and validity issues, its formative or summative purpose, students’ worries and confidence issues, acceptability and level of participation, domains to be assessed, and so forth.

Finally, on implementing this knowledge, and taking Watling et al. (2013) and their study on medical and music learners as a model, we would be pleased to see our students aim for an ever-better performance, learn teaching and self-assessment skills, but recognise and seek life-long external feedback.
Chapter 2

Teaching and Assessment Approaches in Dental Education

Figure 2-1 Flowchart of the study’s research chapters. The second current chapter is highlighted to help in providing an overall view.

2.1 Introduction

Dental education requires the integration and broad understanding of healthcare and basic and oral sciences for students to develop their knowledge, motor and intellectual skills, as well as attitudes and behaviours (Entwistle and Hounsell, 1977; Oliver et al., 2008). For dental education to be successful, particularly in relation to developing habits that encourage a self-directed learning culture (Manogue et al., 2011), it requires a wide range of teaching methodologies as well as an integrated assessment strategy.

Dental schools in the United Kingdom comply with learning outcomes determined by the General Dental Council (GDC) as the regulatory body. The GDC was created in 1956 (Basker, 2006) to ensure high standards of dental education (UK Government...
“Dentist Act 1984”) at all its stages (General Dental Council, 1997). The initial GDC “Recommendations concerning the Dental Curriculum” (1990) evolved into a sequence of three publications entitled “The First Five Years” (1997; 2002; 2008) with increasing detail regarding the expected competency profile of a new graduate. In January 2012 the GDC published its latest guidelines for course providers and awarding bodies entitled “Preparing for Practice: Dental team learning outcomes for registration” (2012a). This document sets out the current learning outcomes, grouped into four integrated and mutually supported domains (Figure 2-2), that a candidate must be able to demonstrate by the time of graduation.

![Figure 2-2 “Preparing for Practice: Dental team learning outcomes for registration” four integrated and mutually supported domains (2012a p. 9).](image)

These learning outcomes reflect the already required “knowledge, skills and attitudes” outlined in the two last versions of The First Five Years (2002; 2008), with the addition of “behaviours” with the aim of developing a rounded professional who practises safely, effectively and professionally (General Dental Council, 2012a). Whilst the 2002 and 2008 versions identified three levels to a developmental framework for each learning outcome (“be competent at”, “have knowledge of” and “be familiar with”), these have now been replaced by a different analytical framework (Pangaro and ten Cate, 2013). This consists of a list of verbs called “keywords” associated with the learning outcomes required levels of Knowledge, Skills and Attitudes/Behaviours (Table 2-1).
Regarding assessment, the GDC “Preparing for Practice” (2012a) document states that students should be assessed rigorously, appropriately and reliably in a demonstrable and clear way against all outcomes. However, it only mentions some assessment methods such as workplace based assessment, portfolios, projects and examinations in passing. Later that year the GDC published the “GDC Standards for Education” in which the third of its four standards concentrated on “Student Assessment” (General Dental Council, 2012c pp. 6-8). Whilst it does not restrict or impose any assessment tool, leaving this to each institution to decide, it does mention the need for ‘reliable’ and ‘valid’ assessment methods to demonstrate achievement of the GDC learning outcomes. Indeed, it sets two important and useful requirements relevant to this study:

- Requirement N. 20: “providers should seek to improve student performance by encouraging reflection and by providing feedback” which, among other proofs, means the need for training in reflection and receiving feedback, and some evidence of reflection.
- Requirement N. 25: asks for valid and reliable “multiple samples of performance” and evidence of “continuous assessment” (General Dental Council, 2012c).

Finally, the GDC regulatory document entitled “Quality Assurance Process” (General Dental Council, 2012b) requires from every new programme a mapping of assessment against the “Preparing for Practice” learning outcomes.

Amalgamating these together, there is a clear message from the regulator to dental educational providers to collect evidence and assume the role and responsibility of preparing students to carry out reflective practice and self-directed learning, to adhere to
a life-long learning (General Dental Council, 2012a).

It is not surprising that King’s College London “Learning, Teaching and Assessment Strategy” (King’s College London Academic Services, 2011) states that students should become increasingly creative, critical and analytical and be able to work independently, taking their own decisions with a sound base of values and an understanding of others. The King’s Strategy goes on to assert that assessment is an integral part of the learning process and must be fair, transparent and encourage learning. Hence, by engaging students in the assessment process and offering them feedback in a notion of “assessment for learning” (Schuwirth et al., 2011), it is designed to enable students to continually improve their work.

In view of the above we felt it appropriate to investigate how dental teaching and particularly assessment are conducted using the GDC learning outcomes as the framework.

2.2 Aim

To investigate the different teaching and assessment methodologies used in dental education using King’s College London Dental Institute (KCLDI) as true model.
To investigate assessment practices and strategies used by selected dental teaching institutions around the world to measure students’ progress particularly in relation to the use of peer-assessment.

2.3 Hypotheses

i. KCLDI encompasses the teaching and assessment of knowledge, skills and attitudes/behaviours in all four GDC domains (Clinical, Communication, Professionalism, and Management and Leadership).

ii. Peer-assessment is employed in more than 50% of the surveyed (inter)national dental teaching institutions.

2.4 Materials and Methods
2.4.1 KCLDI teaching and assessment

All 51 undergraduate coursebooks (syllabus) that constitute the Bachelor of Dental Surgery (BDS) 2011/2012 program at KCLDI were collated. A descriptive quantitative analysis of these was undertaken to determine the current teaching and assessment practices.

The “Preparing for Practice” (General Dental Council, 2012a) learning outcomes for dentists were then used as a framework to map every teaching and assessment session from the KCLDI syllabus using the template shown in Table 2-2. The teaching methodologies at each BDS year were then classified in six clusters: Lectures, Tutorials & Workshops, Self-directed Learning, Practicals & Clinical, E-learning and others. Similarly, assessment sessions were grouped by year and then categorized according to Miller’s Pyramid of Professional Competence (Miller, 1990).

Table 2-2 Template used for mapping the 2012 GDC learning outcomes with BDS teaching and assessing methodologies at KCLDI.

<table>
<thead>
<tr>
<th>GDC DOMAIN</th>
<th>OUTCOMES</th>
<th>Teaching</th>
<th>Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Year/module in which the learning outcome is covered</td>
<td>Course in which the learning outcome is covered</td>
<td>How the learning outcome is assessed</td>
</tr>
<tr>
<td>Clinical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Professionalism</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Management and Leadership</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The KCLDI BDS programme learning outcomes were categorised according to the new GDC taxonomy of “knowledge”, skills” and “attitudes/behaviours” keyword verbs (Table 2-1) for each of the four domains (Clinical, Communication, Professionalism, and Management and Leadership).

2.4.2 Assessment practices and strategies online survey

The resulting list of KCLDI assessment methodologies used across the BDS programme were added with other tools from the dental (Manogue et al., 2001; Albino et al., 2008;
Kramer et al., 2009) and medical (Wass et al., 2001; Shumway and Harden, 2003; Norcini and Burch, 2007; Kogan et al., 2009; Swanwick, 2011) education literature to prepare an assessment practices online survey. These assessment tools list were then mapped to the four GDC “Preparing for Practice” domains (Clinical, Communication, Professionalism, Management and Leadership). To enhance the survey content validity (Streiner and Norman, 2008), the resultant mapping was reviewed by seven internationally renowned dental educators (all members of the Association for Dental Education in Europe (ADEE)), at the “Assessment of Clinical Competencies in Dental Education” Special Interest Group (SIG).

The outcome of this review was presented as an English language online survey (Table 2-3) and distributed online using Survey Monkey® (Palo Alto, California, USA) to selected dental teaching institutions to assess their assessment strategies for measuring students’ progress within each of the four GDC domains. The selected institutions were Dental Schools, Institutes or Centres involved in research publications in the European Journal of Dental Education and/or the Journal of Dental Education between the years 2009 to 2012 with an English website.

In total, 49 dental teaching institutions, representing 15 countries from all 4 continents were selected and their Dean or Director of Education sent an electronic invitation explaining the aims of the study inviting them to consent and complete the online survey. They were asked to identify which of the assessment methods in the list were the most important sources of information to routinely assess students' competence and performance (not seldom-used techniques) at their Dental School / Institute / Centre. They were also requested to choose whether the assessment method was used formatively, summatively or as a hurdle (gateway).

All collected data were entered into an Excel (Microsoft Inc., Seattle, WA, USA) spreadsheet and descriptively analysed using the Statistical Package for the Social Sciences® Windows® version 21 (SPSS Inc. IBM, Chicago, IL, USA).

1 USA dental teaching institutions were excluded as they had recently been surveyed in a similar study Albino, J., Young, S., Neumann, L., Kramer, G., Andrieu, S., Henson, L., Horn, B. & Hendricson, W. 2008. Assessing Dental Students’ Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education. Journal of Dental Education, 72, 1405-1435.
Table 2-3 Survey designed to assess how dental teaching institutions assessed each of the four GDC “Preparing for Practice” domains.

<table>
<thead>
<tr>
<th>Assessment method</th>
<th>Clinical</th>
<th>Communication</th>
<th>Professionalism</th>
<th>Management and Leadership</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multiple choice questions linked to cases</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Multiple choice questions not linked to cases</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Short Answer Questions (complete the sentence), quiz, brief tests and questionnaires</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Essays</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Oral examinations with checklist</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Oral examinations without checklist</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Research projects</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Poster or other presentations</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Observation of students’ performance of technical skills in preclinical labs</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Observation of students’ performance in computer-based or other types of simulations</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Chart-stimulated oral exams (clinical reasoning examinations)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Objective Structured Clinical Examinations (OSCE)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Structured Clinical Operative Test (SCOT)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Triple Jump (1st: Patient interview &amp; examination observation, 2nd: Write assessment and treatment plan, 3rd: Oral exam on examined patient)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) termly</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) every patient</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) occasionally</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Students self-assessment</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Students peer-assessment</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Assessment by patients</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Portfolios</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Learning diaries, log books, record review</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Quantity / Number of procedural units performed by students</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>• Other Assessment Methods (Please specify which ones)</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
2.5 Results

2.5.1 KCLDI teaching and assessment

The complete transcription of the KCLDI syllabus mapped to the GDC learning outcomes framework is included in the Appendices (Appendix 11).

Table 2-4 KCLDI teaching methodologies by course year.

<table>
<thead>
<tr>
<th>Course Year</th>
<th>Lectures</th>
<th>Tutorials &amp; Workshops</th>
<th>Self-directed learning</th>
<th>Practicals &amp; Clinical</th>
<th>e-learning</th>
<th>Others *</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS 1</td>
<td>188</td>
<td>65</td>
<td>9</td>
<td>35</td>
<td>7</td>
<td>11</td>
<td>315</td>
</tr>
<tr>
<td>BDS 2</td>
<td>166</td>
<td>89</td>
<td>58</td>
<td>120</td>
<td>1</td>
<td>4</td>
<td>438</td>
</tr>
<tr>
<td>BDS 3</td>
<td>147</td>
<td>9</td>
<td>15</td>
<td>153</td>
<td>48</td>
<td>1</td>
<td>373</td>
</tr>
<tr>
<td>BDS 4</td>
<td>114</td>
<td>20</td>
<td>12</td>
<td>150</td>
<td>10</td>
<td>2</td>
<td>308</td>
</tr>
<tr>
<td>BDS 5</td>
<td>65</td>
<td>16</td>
<td>0</td>
<td>193</td>
<td>6</td>
<td>2</td>
<td>282</td>
</tr>
<tr>
<td>Total</td>
<td>680</td>
<td>199</td>
<td>94</td>
<td>651</td>
<td>72</td>
<td>20</td>
<td>1716</td>
</tr>
</tbody>
</table>

* poster, miscellaneous

The total number of teaching sessions for the five years of the KCLDI BDS programme was 1716 (Table 2-4). The commonest teaching methodology was Lectures (40%), followed by Practicals & Clinical (38%), Tutorials & Workshops (12%), Self-directed learning (5%), e-learning (4%) and others (poster, miscellaneous) (1%) (Figure 2-3).

Figure 2-3 Graphical representation of the percentage of teaching methodologies from all courses together.
With respect to assessment, the five year programme included 23 different assessment methodologies totalling 153 assessment sessions including formative (55%), summative (30%) and hurdle (15%). The most frequently used assessment tools were structured essays (13%), workplace based assessments marked by tutors (13%) (Tutor Mark Assessments (TMA), Structured Clinical Operative Tests (SCOT), Student Performance Indicators, Paediatric Clinical Competency), Single Best Answers (12%), Extended Matching Questions (11%), OSCEs (9%), quiz, brief tests and questionnaires (10%), Short Answer Questions (7%), oral examinations (5%), skills laboratory practicals (5%), case presentations (5%), portfolios (5%), self-assessment (3%), Patient Assessment Questionnaires (1%), and clinical reasoning examinations (1%).

Remarkably, these assessment methods took different roles as formative, summative or hurdle. Thereby, from the total of 153 assessment sessions, those with a formative purpose were represented mostly by quiz, brief tests and questionnaires, Single Best Answers, workplace-based assessments, Extended Matching Questions, OSCE, and Portfolios. Differently, summative assessment methods included mainly Essays, Short Answer Questions, Single Best Answers, and Extended Matching Questions. Finally, Hurdle intentioned assessments included predominantly workplace-based Tutor Mark Assessments, Orals, and case presentations.

Figure 2-4 Overall assessment framework within the four levels of Miller’s Pyramid (Miller, 1990).
When the assessment methods from the framework were classified using the four levels of Miller’s Pyramid of Professional Competence (Miller, 1990), it was possible to appreciate (Figure 2-4) that the “Knows” level represented the predominant (38.6%) category followed by “Knows How” (33.3%), “Shows How” (13.1%) and “Does” (15.0%).

The overall picture however, changed dramatically when these assessment methodologies were divided according to their purpose (formative, summative or hurdle) (Figure 2-5). Particularly interesting is the low 1% of workplace assessment in relation to the summative group. This might denote the previously mentioned difficulties of clinical instructors to reliably assess learners’ specific struggles (Audétat et al., 2013) as well as clinical and reasoning skills, while at the same time ensuring high-quality patient care (Bowen, 2006; Norcini and Burch, 2007).

![Figure 2-5 Summative, Formative and Hurdle assessment approaches mapped to the four levels of Miller’s Pyramid (Miller, 1990).](image)

Additionally, the distribution of these assessment types varied across the different BDS years (Figure 2-6). Accordingly, the test of factual recognition – Knows – through for example MCQs, together with the higher level assessment of clinical context applications – Knows How – through for example SBA and Essays, represented the majority of all assessment sessions throughout the programme. However, both these groups of assessment methods showed a clear decrease in their proportional representation from BDS 1 (45.5% for ‘Knows’ and 40.9% for ‘Knows How’) to BDS 5 (33.5% for ‘Knows’ and 28.6% for ‘Knows How’). Interestingly, and though the fraction of assessment methods at the ‘Shows How’ level remained below the 17% all
over, it represented a vital stage for students to demonstrate their capacity to reflect and integrate all their professional skills through, for example the Clinical Reasoning Examination and the Clinical Case Presentation. This however paved the way for the higher ‘Does’ level of workplace assessment methods to overtake the ‘Shows How’ level already at BDS 3 (14.7% versus 11.8%, respectively), reaching a higher 21.9% and 21.4% at BDS 4 and BDS 5, respectively.

Figure 2-6 Percentage of assessment sessions grouped according to Miller’s pyramid of professional competence, per year.

The “Preparing for Practice” (General Dental Council, 2012a) guidelines for dentists contain a total of 157 learning outcomes, of which 7 are overarching while 150 are domain specific (95 Clinical, 21 Communication, 19 Professionalism, and 22 Management and Leadership). The KCLDI BDS programme accounted for 142 (90%) of these learning outcomes, which means 15 were not formally covered: 5 in Clinical, 2 in Communication, 4 in Professionalism, and 4 in Management and Leadership domains (Appendix 11 Transcription of the KCLDI syllabus). However, the programme also included a number of 71 teaching sessions in the Clinical domain that did not match any of the 2012 “Preparing for Practice” learning outcomes (Appendix 11 Transcription of the KCLDI syllabus). One should keep in mind that the studied KCLDI 2011/2012 academic programme was developed to comply with the 2002 GDC guidelines (General Dental Council, 2002). It is expected that missing learning outcomes from the 2012 “Preparing for Practice” guidelines are to be gradually introduced in the subsequent
academic years and certainly with the new KCLDI curriculum introduced in 2013.

Coming back to the studied KCLDI BDS programme, the vast majority of the covered learning outcomes corresponded to the Clinical domain (67.4%), followed by Management and Leadership (13.7%), Professionalism (12.0%) and Communication (6.9%). According to the GDC taxonomy, 53.6% of these learning outcomes were classified in the “knowledge” category, whilst 32.2% corresponded to “skills” and 14.2% to “attitudes/behaviours”. These figures vary considerably when outcomes were analysed by domain (Table 2-5). Further, every single learning outcome was aligned with a teaching and assessment methodology as described in Appendix 11.

Table 2-5 Learning outcomes range according to the keyword verbs from the proposed taxonomy for each of the four domains.

<table>
<thead>
<tr>
<th>Learning Outcome Taxonomy</th>
<th>KCLDI learning outcomes according to the GDC domains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Clinical</td>
</tr>
<tr>
<td>Knowledge</td>
<td>%</td>
</tr>
<tr>
<td>Skills</td>
<td>58.6</td>
</tr>
<tr>
<td>Attitudes / Behaviours</td>
<td>3.2</td>
</tr>
<tr>
<td>Total</td>
<td>67.4</td>
</tr>
</tbody>
</table>

2.5.2 Assessment practices and strategies online survey

The assessment survey was answered by 38/49 (78%) of the selected dental schools / institutes / centres within three weeks. The results (Table 2-6) demonstrate that all listed assessment methods were used by at least one surveyed institution, and reveals the wide variety of assessment strategies utilised to assess different domains.

The most widespread assessment method was the OSCE used for the majority (52.6%) of the institutions for the assessment of clinical and communication skills. Subsequently, other commonly utilised methods for the clinical domain were the MCQ linked to cases (39.5%), observation of students' performance of technical skills in preclinical labs (31.6%), structured workplace-based assessment of students' performance (WPBA) with every patient (31.6%), short answer questions (28.9%), and both essays and self-assessment which reported the same usage (26.3%).
Table 2-6 Survey results showing the % of dental teaching institutions using the listed assessment methods for each of the four GDC “Preparing for Practice” domains.

<table>
<thead>
<tr>
<th>Assessment method</th>
<th>Clinical</th>
<th>Communication</th>
<th>Professionalism</th>
<th>Management and Leadership</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Multiple choice questions linked to cases</td>
<td>39.5%</td>
<td>15.8%</td>
<td>15.8%</td>
<td>13.2%</td>
</tr>
<tr>
<td>• Multiple choice questions not linked to cases</td>
<td>21.1%</td>
<td>2.6%</td>
<td>5.3%</td>
<td>7.9%</td>
</tr>
<tr>
<td>• Short Answer Questions (complete the sentence), quiz, brief tests and questionnaires</td>
<td>28.9%</td>
<td>13.2%</td>
<td>13.2%</td>
<td>10.5%</td>
</tr>
<tr>
<td>• Essays</td>
<td>26.3%</td>
<td>13.2%</td>
<td>13.2%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Oral examinations with checklist</td>
<td>21.1%</td>
<td>10.5%</td>
<td>10.5%</td>
<td>5.3%</td>
</tr>
<tr>
<td>• Oral examinations without checklist</td>
<td>21.1%</td>
<td>5.3%</td>
<td>5.3%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Research projects</td>
<td>18.4%</td>
<td>5.3%</td>
<td>5.3%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Poster or other presentations</td>
<td>13.2%</td>
<td>15.8%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>• Observation of students’ performance of technical skills in preclinical labs</td>
<td>31.6%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>• Observation of students’ performance in computer-based or other types of simulations</td>
<td>15.8%</td>
<td>2.6%</td>
<td>2.6%</td>
<td>0%</td>
</tr>
<tr>
<td>• Chart-stimulated oral exams (clinical reasoning examinations)</td>
<td>10.5%</td>
<td>2.6%</td>
<td>0%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Objective Structured Clinical Examinations (OSCE)</td>
<td>52.6%</td>
<td>52.6%</td>
<td>18.4%</td>
<td>15.8%</td>
</tr>
<tr>
<td>• Structured Clinical Operative Test (SCOT)</td>
<td>18.4%</td>
<td>2.6%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>• Triple Jump</td>
<td>7.9%</td>
<td>5.3%</td>
<td>2.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) termly</td>
<td>15.8%</td>
<td>7.9%</td>
<td>10.5%</td>
<td>7.9%</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) every patient</td>
<td>31.6%</td>
<td>21.1%</td>
<td>18.4%</td>
<td>15.8%</td>
</tr>
<tr>
<td>• Structured Workplace-Based Assessment of students’ performance (WPBA) occasionally</td>
<td>21.1%</td>
<td>15.8%</td>
<td>13.2%</td>
<td>15.8%</td>
</tr>
<tr>
<td>• Students self-assessment</td>
<td>26.3%</td>
<td>18.4%</td>
<td>15.8%</td>
<td>13.2%</td>
</tr>
<tr>
<td>• Students peer-assessment</td>
<td>10.5%</td>
<td>10.5%</td>
<td>7.9%</td>
<td>7.9%</td>
</tr>
<tr>
<td>• Assessment by patients</td>
<td>7.9%</td>
<td>7.9%</td>
<td>2.6%</td>
<td>2.6%</td>
</tr>
<tr>
<td>• Portfolios</td>
<td>21.1%</td>
<td>13.2%</td>
<td>13.2%</td>
<td>13.2%</td>
</tr>
<tr>
<td>• Learning diaries, log books, record review</td>
<td>18.4%</td>
<td>5.3%</td>
<td>10.5%</td>
<td>7.9%</td>
</tr>
<tr>
<td>• Quantity / Number of procedural units performed by students</td>
<td>15.8%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>• Other Assessment Methods</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Long case using virtual patients</td>
<td>2.6%</td>
<td>2.6%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Comprehensive incident report</td>
<td>0%</td>
<td>2.6%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Remarkably, the latter was also frequently used to assess each ones’ communication skills (18.4%) and was third after OSCE (52.6%) and WPBA (21.1%) in this domain.
Regarding professionalism, the most frequent practices were WPBA with every patient and OSCE, both being used by the same number of institution (18.4%).

Management and leadership was equally assessed (15.8%) by OSCE, WPBA with every patient, and occasional WPBA. The least frequently used tools from the proposed list were the triple jump (2.6%) and assessment performed by patients (2.6%). Further, only two assessment methods not contained in the survey list were reported by one institution each: Long case using virtual patients for clinical and communication skills (2.6%), and Comprehensive incident report for communication skills (2.6%).

2.6 Discussion

Collating the different teaching and assessment approaches used at KCLDI required a close scrutiny of all the programme coursebooks which provided an excellent opportunity to understand how the curriculum was both organised and constructively aligned (Biggs and Tang, 2011).

The KCLDI BDS programme embraces a wide variety of teaching and assessment methodologies of knowledge, skills and attitudes/behaviours learning outcomes across all four GDC defined domains. On this basis, the first hypothesis of this Chapter can be accepted: KCLDI encompass the teaching and assessment of knowledge, skills and attitudes/behaviours in all four GDC domains (Clinical, Communication, Professionalism, and Management and Leadership).

Some other more specific findings from this Chapter helped to inform the studies in the later sections. An example of this is the balance between “lectures” (40%) and “practicals and clinical” (38%) teaching sessions (Figure 2-3). Lectures are intended to develop student’s knowledge and are mostly utilised at the beginning of the programme (Table 2-4) and are supported by a larger number of tutorial and workshop sessions aimed at increasing students’ understanding of the given knowledge (King's College London Dental Institute, 2012). The subsequent practical and clinical sessions, intended to expand students’ dexterity and clinical skills, are logically concentrated towards the end of the programme.

Both “King’s Learning, Teaching and Assessment Strategy” (King's College London Academic Services, 2011) and the GDC “Preparing for Practice” (General Dental Council, 2012a) guidelines, task dental teaching institutions to prepare increasingly
creative, critical, analytical, reflective and self-directed learner students. Given this, the reported 5% “self-directed learning” teaching sessions might be seen as lower than expected or perhaps desirable and it may be sensible to look to increase this element to facilitate further development of the students higher orders thinking skills (Spronken-Smith and Walker, 2010; Speyer et al., 2011). Similarly, the introduction of assessment tools that promote reflection, critical thinking and continued learning like self- and/or peer-assessment and portfolios, could be considered (Manogue et al., 2011).

A review of the learning outcomes that shape the KCLDI 2011/2012 BDS programme against the new GDC taxonomy (Table 2-1), shows an uneven combination of “Knowledge” (53.6%), Skills (32.2%) and Attitudes/Behaviours (14.2%) (Table 2-5). However and debatably, 58.6% of the Clinical learning outcomes represent knowledge (58.6%) followed by skills (38.2%) and to a much lesser extend Attitudes/Behaviours (3.2%). One would expect this Clinical domain to be a balanced combination of knowledge and skills. However, as mentioned above, the studied KCLDI 2011/2012 BDS academic programme was developed to comply with the 2002 GDC guidelines (General Dental Council, 2002) and does not represent the planned new curriculum introduced in 2013.

The distribution of KCLDI assessment methodologies within Miller’s model shows the robust base of knowledge at the “Knows” (38.6%) and “Knows How” (33.3%) levels which the Institute intends to provide its students with (King's College London Dental Institute, 2012). This is particularly true at the beginning of the programme (Figure 2-6), and allows them to develop and practice their clinical skills firstly in a simulation laboratory environment before progressing to the real workplace of the patients’ clinic. The new 2012 GDC requirements to introduce the teaching and assessment of “behaviours” besides knowledge, skills and attitudes (General Dental Council, 2012a), will certainly create the need to develop and introduce new assessment methodologies which will address the two apical levels of Miller’s model.

Within the 153 KCLDI assessment sessions, formative in-course assessment represented 55% of the total. This, as discussed in Chapter 1 (Assessment purposes in page 43), is in keeping with educational philosophies where assessment for learning should take priority over assessment of learning (Schuwirth and van der Vleuten, 2011b). The use of twenty-three assessment methodologies throughout the programme mean that each competence is assessed in more than one way and that each
methodology is also used to assess more than one competence which enables triangulation, in keeping with current best practice (Manogue et al., 2002; Schuwirth and van der Vleuten, 2011b). The number and nature of assessment methodologies are in accordance with the reported twenty-one assessment methods utilised at the Bachelor of Oral Health program at Griffith University School of Dentistry (Nulty et al., 2010) and the seventeen assessment strategies described to be used by fifty-three U.S. dental schools (Albino et al., 2008), which further supports the process.

A striking feature of the assessment practices survey is how similar the assessment strategies are of dental teaching institutions from different countries and continents. This might be interpreted as a clear sign of the growing international adoption and convergence of higher education practices and models at work today (Collini, 2012 pages 14-15). Additionally, the fact that all the assessment methods listed in the survey were used by at least one institution, together with the point that only two other methods were reported, subscribe to the content validity of the survey itself (Streiner and Norman, 2008).

The most common assessment method reported by the 38 surveyed institutions was the OSCE (52.6%). In contrast, OSCE only represented a 3% among other 17 assessment method used by 53 surveyed US dental schools (Albino et al., 2008). This despite the fact that this multi-competency examination was introduced in Scotland by Harden et al. (1975) almost 40 years ago. Possible explanations for the low usage of the OSCE in American dental schools might be the fact that the US survey was undertaken over 6 years ago or perhaps that the assessment of “Shows how” third level of Miller’s pyramid framework (Miller, 1990), is undertaken through other methods e.g. laboratory practical’s (8%), self-assessment (7%), unit requirements (3%), computer-based simulations (3%) and chart-stimulated evaluation (2%).

When combining the traditional multiple choice questions linked and not-linked to cases in the clinical domain, both our results and those of Albino et al. (2008) show these to be the most commonly used assessment technique (61% and 45%, respectively). This provides further evidence of the current international convergence of higher education practices and models in relation to assessment of the “knows” (factual recall) and “knows how” (apply biomedical information) levels of competence in Miller’s pyramid (Miller, 1990; Wass et al., 2001; Shumway and Harden, 2003).

The fact that in the current study only 10.5% of the surveyed dental teaching institutions
reported the use of peer-assessment for the clinical and communication domains and a lower 7.9% for the professionalism, and Management and leadership domains, denotes that the second hypothesis of the study must be rejected: Peer-assessment is employed in more than 50% of the surveyed (inter)national dental teaching institutions.

However, this proportionally low usage of peer-assessment should not be taken negatively as it is eminently achievable and provides an opportunity to foster our students’ achievements (Reported benefits of peer-assessment in page 54).

Given the role and responsibility of dental education providers to prepare students to be creative, critical and analytical, and to carry out reflective practice and self-directed learning so that they become life-long learners (King's College London Academic Services, 2011; General Dental Council, 2012a), the number of assessment methods that adhere to the American Dental Education Association (ADEA) suggestions to assess critical thinking and problem solving are small. Thus, according to Kramer et al. (2009) the preferred assessment method for these competencies is the “structured essays”, whilst “structured observation” is felt to be only acceptable. Our findings support the idea that only a few assessment methods are related to students’ higher order thinking skills as in the qualitative phase of the SOLO taxonomy (Biggs and Tang, 2011 pp. 90-91). In our experience, these are essays and clinical reasoning examinations (Chapter 4), and peer-assessment (Chapter 5).

On this subject of encouraging students’ reflection, it is important to consider again the work of Biggs and Tang in that “self- and peer-assessment are particularly helpful for training students to reflect on the quality of their own work” (Biggs and Tang, 2011 p. 196). These indications and the positive effects they have on students’ learning have also been highlighted by other education scholars such as Dochy et al. (1999) and Boud and Falchikov (2007). Thus, the wide spread use of self-assessment in all four domains reported in the present study survey is encouraging. Further, although to a lesser degree, some institutions reported the use of peer-assessment (Table 2-6). In contrast, US dental schools in one study did not report the use of peer-assessment and they only used self-assessment (Albino et al., 2008). However, a second study one year later (Haden et al., 2010) on “Curriculum Change” which surveyed 50 US and 5 Canadian dental schools, reported the “clinical student-faculty group practice teams” to mimic the real working environment, as the most frequently mentioned completed (64%) and in progress (19%) innovation in how pre-clinical and clinical dental education was currently being
delivered. Putting both these American studies together (Albino et al., 2008; Haden et al., 2010), and given the potential of a combined peer-assessment followed by self-assessment / reflection strategy (Light et al., 2009 pp. 221-223) (Chapter 1, page 54), it could be interpreted as a missing opportunity to foster students’ academic and reflective skills.

2.7 Conclusion

This Chapter details the variety of teaching and assessment methodologies used at KCLDI and in a wider dental community of 38 institutions from around the world. It also provides insight and describes how a five-year dental curriculum is not only theoretically organised but also delivered.

The findings demonstrate that in addition to traditional assessment methods, a number of newly introduced clinical assessment techniques, more frequently described in the medical literature (Swanwick, 2011), are also being employed in undergraduate dental education. Some of these have good education evidence of their benefit to future learning (Topping, 1998; Boud and Falchikov, 2007 pp. 114-143; Biggs and Tang, 2011 pp. 217, 245-246, 266-267). One example of this is peer-assessment although some studies limit its use to communication and interpersonal skills and health promotion competencies, highlighting especially its limits by the peers’ level of knowledge and ability (Arnold, 2002; Sluijsmans et al., 2002; Nilson, 2003; Norcini, 2003a). However, the evaluation detailed in the following chapters challenges this position and demonstrates that undergraduate dental students can effectively peer-assess each other in all four GDC domains: Clinical, Communication, Professionalism, and Management and Leadership.
Chapter 3

Developing and Piloting a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback

Figure 3-1 Flowchart of the study’s research chapters. The third chapter is highlighted to help in providing an overall view.

3.1 Introduction

As described in Chapter 1, during the last decade or so, the use of peer-assessment in dental education has been encouraged (page 52). Consequently, recent studies on dental students have reported different peer-assessment methods, both in pre-clinical and clinical settings, with the purpose of evaluating mainly its reliability.

From the studies discussed in Chapter 1, it is known that peer-assessment scores, as compared to experienced staff marks, are reliable when pre-clinical students assess each
others’ technical skills (Satterthwaite and Grey, 2008; Karl et al., 2011; Taylor et al., 2013), and simulated clinical interprofessional communication and examination skills (Ogden et al., 2000). In contrast to the latter study, Lanning et al. (2011) found no correlation between peer-and staff scores in communication skills.

Further, and in agreement with the education literature (Topping, 1998), it has also been reported that dental students self-assessment and peer-assessment marks differ widely (Larsen and Jeppe-Jensen, 2008), and that peer scores reflect trainer ones more accurately (Evans et al., 2007).

Notwithstanding the contribution of these studies, there is still much to know about dental students peer-assessment. It is arguable that despite providing detailed information on the marking phase of the peer-assessment procedures, none of the above reports offered any information on the “learning” usefulness of the experience. Furthermore, it is not known whether dental peers can reliably assess one another in other domains of their practice such as knowledge, skills, professionalism, and management.

In consideration of the above and given the previous chapter’s reported low usage of peer-assessment by surveyed dental teaching institutions, it was deemed necessary to primarily develop and pilot (current Chapter), subsequently implement and psychometrically analyse (Chapters 5 and 6) a peer-assessment protocol especially designed for undergraduate dental students peer-assessment at the top two levels of Miller’s pyramid (Miller, 1990), that could take advantage of a three-pronged strategy. That is, the enhanced potential of a combined prospective peer-assessment and peer-feedback, followed by a self-assessment and reflection phases (Schön, 1983; Ashley et al., 2006; Light et al., 2009 pp. 221-223). All this should contribute to the students’ formative phase of learning (assessment for learning) and help to develop a rounded self-directed, life-long learner and reflective dental practitioner (American Commission on Dental Accreditation CODA, 2010; Australian Dental Council, 2010; General Dental Council, 2012a).

### 3.2 Aim

The aim of the study in this Chapter was to develop and pilot a structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students’ skills in all four clinical, communication, professionalism, and management
and leadership GDC domains, which would be used as an informed framework for the provision of immediate peer-feedback.

3.3 Hypotheses:

i. The developed prospective peer-assessment protocol reliably allows pre-clinical dental students to identify competence differences in various domains.

ii. The developed prospective peer-assessment protocol reliably allows clinical dental students to identify performance differences in various domains.

iii. Most students have positive perceptions of the peer-assessment and peer-feedback protocol.

3.4 Materials and Methods

3.4.1 Developing the instruments

Two widely used workplace-based assessment forms designed for the trainer assessment of medical/dental students and trainees behavioural levels at the top of Miller’s pyramid (1990) in a naturalistic setting (Boursicot et al., 2011), namely Direct Observation of Procedural Skills (DOPS) (Appendix 6) and the Mini-Clinical Evaluation eXercise (Mini-CEX) (Appendix 5) (Norcini and Burch, 2007; The Royal College of Surgeons of England, 2014), were used as starting point templates to develop three new peer-assessment instruments.

The reason behind choosing DOPS and Mini-CEX as inspiring frameworks was based on the aims of the study of developing peer-assessment tools which allowed students to peer-assess and provide face-to-face immediate dialogic peer-feedback in all clinical, communication, professionalism, and management and leadership GDC domains. The mini-Peer Assessment Tool (mini-PAT) (Appendix 10) was discarded as it is intended for anonymous peer-assessment of students’ overall humanistic performance (Archer et al., 2008; Norcini, 2011). The newly developed forms were designed as a framework for a continuous and structured peer-assessment and peer-feedback protocol of undergraduate dental students’ specific procedures pre-clinical competence and clinical performance. A pre-clinical peer-DOPS for peer-assessment of any training procedure performed at the simulation skills laboratory, along with a clinical peer-DOPS for the
peer-assessment of whichever clinical procedure students performed on their patients, and a clinical peer-mCEX for the peer-assessment of first time patients’ clinical examinations/assessment (new patients coming to the clinic), were developed.

Both pre-clinical and clinical peer-DOPS and the clinical peer-mCEX forms included a new general layout, which initially kept the traditional norm-referenced assessment scale of Below expectation, Borderline, Meets expectations and Above expectations (Norcini, 2011) as shown in Figure 3-2.

![Dental Institute](image_url)

**Dental Institute**

**Peer-Assessment & Peer-Feedback**

**Direct Observations of Procedural Skills (DOPS)**

*(Please complete this Form using a cross (X) and CAPITAL LETTERS)*

I. Assessor student’s name: ______________________________________________________

II. Assessor's KCL N° ____________________________________________________________

III. Date: ________________________________________________________________

IV. Have you received training as assessor in Workplace Based Assessment: □ Yes □ No

V. How many times have you assessed previous DOPS with any trainee: 0 1 2-5 6-9 >9

VI. Trainee student’s name: ____________________________________________________

VII. Trainee’s KCL N° ____________________________ Trainee’s role: □ Assistant □ Dentist

VIII. Number of times this procedure has been performed by trainee: 0 1 2-5 6-9 >9

IX. Case Complexity: □ Low □ Moderate □ High  X. Time assessor observing: ____________ minutes.

<table>
<thead>
<tr>
<th>XI. Please grade the following areas using the given scale:</th>
<th>Below expectation</th>
<th>Borderline</th>
<th>Meets expectations</th>
<th>Above expectations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

*Figure 3-2 Initial header of the peer-assessment DOPS form which maintained the traditional norm-referenced assessment scale.*

However, while preparing the training session to be delivered to students (explained below), it was soon noticed this scale, designed for trainer assessment of trainees, would have been difficult for undergraduate students to use as it was understood they might not have the experience/knowledge to judge where their peers’ work was in relation to the “norm”. In the same way, Norcini (2003a) had already reported this norm-referenced scale might prove difficult for junior doctors to make judgements about their peers’ performance quality. Consequently, taking the example of The Royal College of Surgeons of England (The Royal College of Surgeons of England, 2014), a criterion-referenced scale containing four written descriptions based on the needed frequency of clarification, replaced the norm-referenced scale (Figure 3-3). An “unable to comment” option when a given behaviour was not observed was also included.
In order to contribute to the content validity of each of the three new instruments, the most relevant learning outcomes from the respective pre-clinical and clinical courses were selected to represent the corresponding assessment scales (Streiner and Norman, 2008). Based on blueprinting principles (Crossley et al., 2002b), pre-clinical peer-DOPS scale represented the main learning outcomes of the BDS 2 coursebooks (Table 3-1). Similarly, clinical peer-DOPS (Table 3-2) and peer-mCEX (Table 3-3) forms represented both BDS 5 coursebooks learning outcomes, as well as those learning outcomes students must be able to demonstrate by the end of their training in order to register with the GDC as a dental professional (General Dental Council, 2012a).

After several drafts, five internal pre-clinical and clinical teachers (each of whom had at least seven years of teaching experience) ensured both pre-clinical and clinical peer-DOPS as well as clinical peer-mCEX forms sampled all the relevant and pertinent domains and appeared appropriate for the intended purpose (Streiner and Norman, 2008).
Table 3-1 Pre-clinical peer-DOPS checking of the content validity. Each item to be assessed was present in one or more BDS 2 coursebooks learning outcomes. The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

<table>
<thead>
<tr>
<th>Item</th>
<th>Peer-DOPS assessment items for BDS 2</th>
<th>Learning Outcomes KCLDI coursebook</th>
<th>GDC domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge of technique and dental materials used for this procedure</td>
<td>- Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Clinical</td>
</tr>
<tr>
<td>2</td>
<td>Preparing for procedure according to taught protocol</td>
<td>- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Clinical</td>
</tr>
<tr>
<td>3</td>
<td>Technical skills, manual dexterity and instruments handling</td>
<td>- Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Clinical</td>
</tr>
<tr>
<td>4</td>
<td>Working position and indirect vision</td>
<td>- Conservative Dentistry Pre-Clinical and Clinical - Periodontology</td>
<td>Clinical</td>
</tr>
<tr>
<td>5</td>
<td>Following sequence and completing accurately all steps of the procedure</td>
<td>- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Clinical</td>
</tr>
<tr>
<td>6</td>
<td>Observing aseptic technique/ Infection control and safe use of instruments</td>
<td>- Dental Institute Infection Control - Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical - Periodontology</td>
<td>Clinical</td>
</tr>
<tr>
<td>7</td>
<td>Seeking help where appropriate</td>
<td>- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Professionalism</td>
</tr>
<tr>
<td>8</td>
<td>Managing time/punctuality effectively</td>
<td>- Integrated Clinical Care - Conservative Dentistry Pre-Clinical and Clinical</td>
<td>Management &amp; leadership</td>
</tr>
<tr>
<td>9</td>
<td>Supporting and communicating effectively with colleagues and tutors</td>
<td>- Integrated Clinical Care - Communication in Dentistry</td>
<td>Communication</td>
</tr>
<tr>
<td>10</td>
<td>Overall ability to perform procedure</td>
<td>- Integrated Clinical Care - Communication in Dentistry</td>
<td>Clinical</td>
</tr>
<tr>
<td>11</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>- Conservative Dentistry Pre-Clinical and Clinical - Periodontology</td>
<td>Professionalism</td>
</tr>
</tbody>
</table>

Eventually, pre-clinical peer-DOPS forms contained 11 non-compounded items (Mackillop et al., 2011b), designed for the purpose of peer-assessment of any training procedure performed at the simulation skills laboratory, while clinical peer-DOPS ones contained 13 non-compounded items. Similar to the latter one, peer-mCEX forms contained 9 non-compounded items, and were both intended for peer-assessment of whichever clinical assessment or procedure students performed on their patients. Both clinical peer-DOPS and peer-mCEX forms’ items resembled very much the original forms (Norcini and Burch, 2007) and could be used by any clinical BDS student independent of her/his level of training.
Table 3-2. Clinical peer-DOPS checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the “Preparing for Practice” GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

<table>
<thead>
<tr>
<th>Item</th>
<th>Peer-DOPS assessment items for BDS 5</th>
<th>Learning Outcomes</th>
<th>GDC 2012</th>
<th>GDC domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demonstrates understanding of indications, dental materials, complications and technique of the procedure</td>
<td>- Primary Dental Care</td>
<td>- 1.1</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 1.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- 1.1.10</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Obtains informed consent after explaining procedure &amp; possible complications</td>
<td>- Primary Dental Care</td>
<td>- 1.5.3</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td>- 1.5.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 3.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- 3.4</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Demonstrates appropriate preparation pre-procedure</td>
<td>- Primary Dental Care</td>
<td>- 1.8.1</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td>- 1.8.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Administers effective analgesia or safe sedation</td>
<td>- Oral Surgery</td>
<td>- 1.7.5</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Sedation</td>
<td>- 1.7.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Therapeutics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Demonstrates appropriate technical ability in line with usual practice</td>
<td>- Primary Dental Care</td>
<td>- 1.9</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td>- 1.11- 1.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Demonstrates aseptic technique/Infection control &amp; safe use of instruments &amp; sharps</td>
<td>- Decontamination Skills Course</td>
<td>- 1.8.2</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Dental Institute Infection Control</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Deals with unexpected events or seeks help when appropriate</td>
<td>- Oral Surgery</td>
<td>- 1.8.4</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 9.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Orthodontics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Completes post procedure managements</td>
<td>- Primary Dental Care</td>
<td>- 1.7.11</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 1.8.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Communication skills (patient &amp; team)</td>
<td>- Communication in Dentistry</td>
<td>- 3.1</td>
<td>Communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 3.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td>- 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Special Care Dentistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Organisation/efficiency and time management</td>
<td>- Primary Dental Care</td>
<td>- 10.2</td>
<td>Management &amp; leadership</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- 10.2</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Consideration of patient/professionalism</td>
<td>- Oral Surgery</td>
<td>- 6</td>
<td>Professionalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Dental Public Health</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Overall ability to perform procedure</td>
<td>- Primary Dental Care</td>
<td>- Overarching outcomes</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Surgery</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>- Oral Surgery</td>
<td>- 9.3</td>
<td>Professionalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Dental Public Health</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
</tbody>
</table>
Table 3-3. Peer-mCEX checking of the content validity. Each item to be peer-assessed was mapped to the BDS 5 coursebooks and the “Preparing for Practice” GDC learning outcomes (2012a). The GDC domains have been coloured according to the four categories described in the same document and graphically represented in Figure 2-2 (page 93).

<table>
<thead>
<tr>
<th>Item</th>
<th>Peer-mCEX assessment items for BDS 5</th>
<th>Learning Outcomes</th>
<th>GDC 2012</th>
<th>GDC domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interviewing/history taking skills</td>
<td>- Paediatric Dentistry</td>
<td>- 1.2.1</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Orthodontics</td>
<td>- 3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Therapeutics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Physical examination skills</td>
<td>- Paediatric Dentistry</td>
<td>- 1.2.2</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Orthodontics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic skills and underlying knowledge base</td>
<td>- Orthodontics</td>
<td>- 1.1.3</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 1.1.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Acute Dental Care</td>
<td>- 1.4.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Radiology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Communication and listening skills</td>
<td>- Communication in Dentistry</td>
<td>- 3,1</td>
<td>Communication</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Paediatric Dentistry</td>
<td>- 3.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- 4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td>- 5</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Special Care Dentistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Clinical judgment and decision making</td>
<td>- Paediatric Dentistry</td>
<td>- Overarching outcomes</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Orthodontics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Radiology</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Oral Disease</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Special Care Dentistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Consideration for patient/professionalism</td>
<td>- Oral Surgery</td>
<td>- 6</td>
<td>Professionalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Special Care Dentistry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>organisation/efficiency and time management</td>
<td>- Primary Dental Care</td>
<td>- 10.2</td>
<td>Management &amp; leadership</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Overall clinical competence</td>
<td>- Oral Surgery</td>
<td>- Overarching outcomes</td>
<td>Clinical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>- Oral Surgery</td>
<td>- 9,3</td>
<td>Professionalism</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Dental Public Health</td>
<td>- Overarching outcomes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Portsmouth integrated team care</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Additionally, with the purpose of stimulating reflection on practice (Brown and Manogue, 2001), all three instruments incorporated an item on “Students’ insight into their performance” to judge the trainees’ self-assessment (Prescott-Clements et al., 2011), which was only to be graded after providing feedback and agreeing Challenges/Actions. Finally, a 6-point Likert scale for students to rate the utility of giving/receiving feedback as a technique to improve their future performance, written
instructions on how to complete the forms and a wider explanation of the grading scale were also included.

Before starting the pilot, BDS Year 2 and Year 5 students were also asked to review the wording and content of all three forms and then use them once before feeding back. This process identified two areas of student concern. The first related to the new criterion-referenced scale and the need to grade the frequency of peer “clarification” while working. This was felt to negatively affect peer-collaboration as they would refrain from asking questions in order to obtain a better assessment. As this was not the intention of the exercise, the criterion-referenced scale was again changed this time presenting the students with two different six-option educationally-referenced scales (Figure 3-4 and Figure 3-5). The decision in favour of these types of scales was taken in the search for an anchor that was more easily understood by junior students, as several studies have reported they improve the acceptability, validity and reliability of the assessment method (Beard, 2011; Crossley et al., 2011; Crossley and Jolly, 2012).

![Figure 3-4 Header of the first option of an educationally-referenced scale with written explanations of progress presented to students.](image-url)
These two late alternatives were much better accepted by students though they asked to merge the two anchor versions together. The resulted combined six-option educationally-referenced contained a similar graphical and shortened written anchor of the desired “increasing ability over time” (Figure 3-6). This growing capability anchor emphasised, for marking purposes, the end-of-year expectations of their respective course of training (Playford et al., 2013), aimed at facilitating even more junior students’ understanding and use of the scale. For example, when a student first performs a practical task, they would be peer-rated as “starting to develop” the ability for that task. Subsequently, she or he would ideally progress to “initial capability”, followed by “constant acceptable”, “constant clear”, “constant good”, and finally “constant extremely good” ability. Consequently, the new anchor did not require students to make any judgements about the quality of performance or frequency of clarification. Further, as suggested by Pangaro and ten Cate (2013), the whole assessment framework was designed to require a reduced cognitive load from “observing” students, in order to improve their rating accuracy.
The second students’ concern was the extensive general information they had to fill in at the beginning of each form. Further, they were uncomfortable with the use of the word “assessor” in relation to the “observing” student, as they were unhappy to appear “assessing” their peers. In view of this, the header of all forms was simplified and all references to “assessor” were replaced by “observer” (Figure 3-6).

Final piloted versions of pre-clinical peer-DOPS, clinical peer-DOPS and clinical peer-mCEX can be found in the Appendixes Book as Appendix 12, Appendix 13, and Appendix 14, respectively.

---

**Figure 3-6** Header of the six-point educationally referenced scale used in both pre-clinical and clinical peer-assessment instruments which asks the “observing” student to judge their peer’s ability over time.

### 3.4.2 Piloting and data collection

In January 2012, 26 invited students (18 females and 8 males, aged 18 to 40, mean=24.3, sd=5.9) comprising two groups, consented to participate in this peer-assessment and peer-feedback pilot study. The first group consisted of 10 pre-clinical BDS 2 Conservative Dentistry students who were under a single clinical supervisor and working at neighbouring phantom heads. The second group comprised 16 clinical BDS 5 Primary Dental Care (PDC) students who worked on the same day of the week as
clinical partners (alternating dentist/assistant roles).

At the start of the study and following suggestions from the education (Sluijsmans et al., 2002) and clinical assessment (Norcini, 2003a; Cook and Beckman, 2006; Hassell et al., 2012) literature, each group of students received a 45-minute peer-assessment training and familiarisation session delivered by the same researcher (JT) relating to observation, peer-assessment, peer-feedback, action plan and completion of the instruments. Using written/video examples and role-playing, they learnt and practiced how to give (observing student) and receive (training student) confidential, brief, constructive, task-focused (Kluger and DeNisi, 1996), process-oriented (Harks et al., 2013), and immediate dialogic feedback (Epstein, 2007; Miller and Archer, 2010; Finn and Garner, 2011), using their peer-assessment form domains as a framework (Beard et al., 2012). BDS 2 students (organised in fixed pairs) working at neighbouring phantom heads and BDS 5 clinical partners (randomly allocated each session), acted as “observer” and “trainee”, respectively during the first half of the day and then switched roles during the second half of the day.

BDS 2 students performed their own procedures as normal while “observing” their peers’ pre-clinical work every 15 minutes to avoid interfering with their own work. BDS 5 students performed their usual clinical activities in pairs, as it is usually done in dental education (Ahmad et al., 2012), so that the assistant student “observed” the dentist student while treating the patient together. The observed procedure was then used to score each of the respective pre-clinical or clinical peer-assessment forms selecting and ticking one of the six options of the educationally-referenced scale (Figure 3-6) for every domain. If a given behaviour was not observed they ticked the “unable to comment” option. These scores provided a grounded framework to engage in an informed dialogic feedback. Subsequently, they agreed an appropriate action plan to address any developmental needs (Topping, 1998; Miller and Archer, 2010). Finally, after signing the forms and placing them in a specially designed delivery box, students self-assessed and reflected based on the received feedback and noted their thoughts in a private reflection diary.

In order to investigate students’ perceptions of the prospective peer-assessment and peer-feedback protocol, during the final session of peer-assessment, both groups anonymously answered the questions presented in Table 3-4, using a 5-point Likert-scale.
Table 3-4 Anonymous questionnaire to study students’ perceptions presented to both BDS 2 and BDS 5 groups at the end of the peer-assessment exercise.

<table>
<thead>
<tr>
<th>To what extent do you agree that the peer-assessment and feedback protocol used in this study:</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Nor Agree nor disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>…assessed areas that correspond to your activity in the pre-clinic/clinic?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…have helped you to identify learning needs and to improve your performance?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>…was acceptable and fair?</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3.4.3 Data analysis

All peer-assessment forms data were manually digitised by the same researcher (JT) into a spreadsheet. To analyse students’ peer-assessment scores, each of the six levels of “increasing ability over time” of the educationally referenced scale (Figure 3-6) was assigned a numerical value from 1 to 6. Thus, the “Starting to develop” initial stage of ability was given a score 1; the “Show initial capability” a score 2 and so on until the highest “Show constant extremely good ability” which was given a score 6. Subsequently, scores were checked for normality assumptions using histogram before carrying out any parametric analysis. The reliability of the tools scores was assessed independently using Generalizability coefficient (Cronbach and Shavelson, 2004).

Descriptive statistics were used to summarise peer-assessment scores, peer-observation time, peer-feedback time, and the utility of giving/receiving feedback. The same method was used to describe students’ perceptions of the studied peer-assessment and peer-feedback protocol. When comparing various measures observed for BDS 2 and BDS 5 groups, independent samples t-test was used.

In order to study whether the developed peer-assessment tools allowed students to identify competence (pre-clinical BDS 2) and performance (clinical BDS 5) differences
in various domains, a one way Analysis of Variance (ANOVA) was carried out to compare the mean scores of each form domains, separately. Where the ANOVA showed significant results, a post-hoc analysis was carried out using Tukey's test. The total peer-assessment scores observed at various time points during the study period were compared using repeated measures ANOVA for BDS 2 and BDS 5 groups, separately.

To evaluate the criterion validity of the peer-assessment instruments, that is, their correlation with another measure of the same outcome considered as the gold standard (Streiner and Norman, 2008) (further described in Chapter 1, section Assessment principles on page 45), a Pearson correlation analysis between the mean BDS 2 and BDS 5 students’ peer-assessment scores were given along the study period and their respective official high stakes end-of-year mean examination mark, was performed.

Further, to investigate a possible educational effect of the prospective peer-assessment exercise on participating students’ academic performance, independent samples t-test was used to compare the high stakes end-of-year mean examination marks of the 26 BDS 2 and BDS 5 students who used the peer-assessment protocol with their 142 and 146 classmates, respectively, who did not take part in the peer-assessment study.

All statistical analyses were carried out using SPSS® version 21 except for the Generalizability coefficient which was calculated using the software EduG 6.1e (Neuchatel, Switzerland).

### 3.5 Results

In line with current best practice (Topping, 1998; Norcini, 2003a), one researcher (JT) carefully organised, delivered and monitored the whole piloting peer-assessment process. Thus, starting on February 2012 and during six fortnightly occasions for BDS 2 and seven for BDS 5, students observed, assessed and provided feedback to one another. BDS 2 students completed 57 pre-clinical peer-DOPS forms (mean=5.7 per student) whilst BDS 5 students carried out 104 clinical peer-DOPS encounters (mean=6.5 per student). Only two peer-mCEX forms were submitted in the delivery box, probably due to the advanced period of the academic year. It should be born in mind that peer-mCEX forms were designed to be used for the clinical examinations and assessment of new patients coming to the clinic. At the time this study started, students were already ending their clinical cases before their examinations, and new patients...
were not regularly seen. Therefore, considering the impracticality of performing any
analysis with these reduced numbers of peer-mCEX forms, it was decided not to include
them in the pilot study. Nonetheless, as presented in Chapter 5 (page 154), peer-mCEX
forms were effectively included in the implementation phase of this study which
covered the whole academic year.

BDS 2 students were assessed with pre-clinical peer-DOPS across seven different pre-
clinical procedures, ranging from composite, amalgam and temporary restorations to
root canal treatments and direct veneers. BDS 5 students were assessed with clinical
peer-DOPS in nineteen clinical procedures, including oral health instruction,
impression, bite and face-bow registration, composite, amalgam and temporary
restorations, crown, bridge and veneer preparation and cementation, root canal
treatments, wax try-in and root surface debridement.

Peer-assessment scores from both groups ranged from 2 (Show initial capability) to 6
(show constant extremely good ability) (mean=5.0, sd=0.7, mode=5), and were
normally distributed. Generalizability coefficients for pre-clinical BDS 2 peer-DOPS
was performed as a crossed three-facet (10 trainee (t) × 6 occasions (o) × 11 items (i))
random-effects (students worked in fixed pairs throughout the study), while for clinical
BDS 5 peer-DOPS a nested three-facet (16 trainee × 7 occasions × 13 items) random-
effects (students worked in random pairs), was used. Thus, the Generalizability
coefficient for BDS 2 was 0.62 for six encounters whereas for BDS 5 it was 0.67 for
seven encounters. The variance analysis for both pre-clinical BDS 2 peer-DOPS and
clinical BDS 5 peer-DOPS tools is shown in Table 3-5.

Table 3-5 Variance components estimates for each peer-assessment tools.

<table>
<thead>
<tr>
<th>Variance components</th>
<th>BDS-2 peer-DOPS Proportion of overall variance</th>
<th>BDS-5 peer-DOPS Proportion of overall variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trainee (t)</td>
<td>20.7%</td>
<td>3.6%</td>
</tr>
<tr>
<td>Occasion (o)</td>
<td>8.7%</td>
<td>5.0%</td>
</tr>
<tr>
<td>Item (i)</td>
<td>4.1%</td>
<td>3.1%</td>
</tr>
<tr>
<td>to</td>
<td>6.2%</td>
<td>6.5%</td>
</tr>
<tr>
<td>ti</td>
<td>16.5%</td>
<td>5.7%</td>
</tr>
<tr>
<td>oi</td>
<td>6.9%</td>
<td>15.8%</td>
</tr>
<tr>
<td>toi</td>
<td>36.9%</td>
<td>60.3%</td>
</tr>
</tbody>
</table>

The overall mean peer-assessment scores for BDS 2 and BDS 5 groups, along with their
respective peer-observation time, peer-feedback time, and utility of giving/receiving
feedback are presented and compared in Table 3-6.
Table 3-6 Mean and standard deviation (sd) of the peer-observation and feedback times (minutes), overall peer-assessment score (scale 1 to 6), and students’ perception of the utility of giving and receiving feedback to improve future performance (scale 1 to 6), for each of the studied groups.

<table>
<thead>
<tr>
<th>Variables</th>
<th>BDS 2</th>
<th>BDS 5</th>
<th>p value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation time</td>
<td>153.2 (28.2)</td>
<td>100.2 (21.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Feedback time</td>
<td>6.7 (2.9)</td>
<td>4.8 (1.9)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Overall peer-assessment score</td>
<td>4.8 (0.8)</td>
<td>5.6 (0.8)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Utility of giving feedback</td>
<td>4.8 (1.0)</td>
<td>5.0 (0.6)</td>
<td>0.34</td>
</tr>
<tr>
<td>Utility of receiving feedback</td>
<td>5.1 (0.8)</td>
<td>5.3 (0.5)</td>
<td>0.19</td>
</tr>
</tbody>
</table>

The mean peer-assessment scores for each of the eleven BDS 2 peer-DOPS domains (Table 3-7) showed significant differences (F=3.94, p<0.0001) between these 11 items, with the post-hoc analysis using Tukey's test revealing the better performed “Observing aseptic technique ...” (item 6) was significantly (p=0.04) different from all other items. Similarly, BDS 5 mean peer-assessment scores for the thirteen peer-DOPS items (Table 3-8) also differed significantly (F=6.55, p<0.0001). Peer’s scores for "Consideration of patients/ professionalism" (item 11) were statistically higher than all other items (p=0.02).

Table 3-7 Mean and standard deviation (sd) mark for each of the eleven pre-clinical peer-DOPS items showing the ability of pre-clinical BDS 2 participating students to identify differences (F=3.94, p<0.0001) in their peers’ performance (57 completed forms).

<table>
<thead>
<tr>
<th>Item</th>
<th>Pre-clinical peer-DOPS assessment items for BDS 2</th>
<th>Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge of technique and dental materials used for this procedure</td>
<td>4.7 (0.7)</td>
</tr>
<tr>
<td>2</td>
<td>Preparing for procedure according to taught protocol</td>
<td>4.7 (0.6)</td>
</tr>
<tr>
<td>3</td>
<td>Technical skills, manual dexterity and instruments handling</td>
<td>4.8 (0.7)</td>
</tr>
<tr>
<td>4</td>
<td>Working position and indirect vision</td>
<td>4.8 (0.7)</td>
</tr>
<tr>
<td>5</td>
<td>Following sequence and completing accurately all steps of the procedure</td>
<td>4.8 (0.6)</td>
</tr>
<tr>
<td>6</td>
<td>Observing aseptic technique/ Infection control and safe use of instruments</td>
<td>5.3 (0.7)</td>
</tr>
<tr>
<td>7</td>
<td>Seeking help where appropriate</td>
<td>4.8 (0.9)</td>
</tr>
<tr>
<td>8</td>
<td>Managing time/punctuality effectively</td>
<td>4.7 (0.8)</td>
</tr>
<tr>
<td>9</td>
<td>Supporting and communicating effectively with colleagues and tutors</td>
<td>5.0 (0.8)</td>
</tr>
<tr>
<td>10</td>
<td>Overall ability to perform procedure</td>
<td>4.8 (0.8)</td>
</tr>
<tr>
<td>11</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>4.5 (0.8)</td>
</tr>
</tbody>
</table>

The prospective peer-DOPS marks of every fortnightly assessment occasion for pre-clinical (BDS 2) and clinical (BDS 5) students are shown in Figure 3-7. The repeated measures ANOVA of total peer-DOPS scores observed at the various time points (occasions), showed that the overall peer-assessment scores differed significantly between occasions (p<0.0001) for both BDS 2 and BDS 5 groups.
Chapter 3 Developing and Piloting a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback

Table 3-8 Mean and standard deviation (sd) mark for each of the thirteen clinical peer-DOPS items showing the ability of clinical BDS 5 participating students to identify differences ($F=6.55, p<0.0001$) in their peers’ performance (104 completed forms).

<table>
<thead>
<tr>
<th>Item</th>
<th>Clinical peer-DOPS assessment items for BDS 5</th>
<th>Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demonstrates understanding of indications, dental materials, complications and technique of the procedure</td>
<td>5.4 (1.0)</td>
</tr>
<tr>
<td>2</td>
<td>Obtains informed consent after explaining procedure &amp; possible complications</td>
<td>5.6 (0.6)</td>
</tr>
<tr>
<td>3</td>
<td>Demonstrates appropriate preparation pre-procedure</td>
<td>5.6 (0.7)</td>
</tr>
<tr>
<td>4</td>
<td>Administers effective analgesia or safe sedation</td>
<td>5.7 (0.6)</td>
</tr>
<tr>
<td>5</td>
<td>Demonstrates appropriate technical ability in line with usual practice</td>
<td>5.5 (0.6)</td>
</tr>
<tr>
<td>6</td>
<td>Demonstrates aseptic technique/Infection control &amp; safe use of instruments &amp; sharps</td>
<td>5.4 (1.3)</td>
</tr>
<tr>
<td>7</td>
<td>Deals with unexpected events or seeks help when appropriate</td>
<td>5.4 (1.0)</td>
</tr>
<tr>
<td>8</td>
<td>Completes post procedure managements</td>
<td>5.3 (0.8)</td>
</tr>
<tr>
<td>9</td>
<td>Communication skills (patient &amp; team)</td>
<td>5.4 (0.8)</td>
</tr>
<tr>
<td>10</td>
<td>Organisation/efficiency and time management</td>
<td>5.7 (0.6)</td>
</tr>
<tr>
<td>11</td>
<td>Consideration of patient/professionalism</td>
<td>6.0 (0.4)</td>
</tr>
<tr>
<td>12</td>
<td>Overall ability to perform procedure</td>
<td>5.7 (0.6)</td>
</tr>
<tr>
<td>13</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>5.7 (0.5)</td>
</tr>
</tbody>
</table>

Figure 3-7 Graphical representation of prospective peer-DOPS marks for every fortnightly assessment occasions (mean and standard deviation).
Regarding criterion validity, mean peer-DOPS scores of the 10 individual BDS 2 participating students showed a positive correlation ($r=0.593$) with their high stakes end-of-year examination marks. However and despite this positive correlation, this was not statistically significant ($p=0.071$). Similarly, for the 16 BDS 5 participating students the correlation with their high stakes end-of-year examination marks was also positive ($r=0.505$) and likewise, it was not statistically significant ($p=0.051$).

Subsequently, to evaluate the possible educational impact of the prospective peer-assessment protocol, all participating students’ mean high stakes end-of-year examination marks were compared to those examination marks of the rest of their respective classes. While BDS 2 high stakes examinations consisted of written short answer questions and short note questions, online multiple choice questions, and a combined clinical skills examination viva and practical, BDS 5 ones involved an essay on clinical scenarios, online short answer questions, short note questions and multiple choice questions, OSCE, clinical reasoning examination, and an oral case presentation. Thus, as shown in Table 3-9, both BDS 2 and BDS 5 studied groups did not show a statistically significant difference in their high stakes end-of-year examination marks compared to their classmates who did not use the peer-assessment protocol.

Table 3-9 Comparison of the BDS 2 and BDS 5 peer-assessment participating students high stakes end-of-year examination marks with those of their respective classmates who did not take part in the study.

<table>
<thead>
<tr>
<th>Group of students</th>
<th>N</th>
<th>Mean end-of-year examination mark (sd)</th>
<th>$p$ value of difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS 2 peer-assessment group</td>
<td>10</td>
<td>60.5 (7.2)</td>
<td>0.886</td>
</tr>
<tr>
<td>BDS 2 All non-participating</td>
<td>142</td>
<td>60.1 (8.5)</td>
<td></td>
</tr>
<tr>
<td>BDS 5 peer-assessment group</td>
<td>16</td>
<td>67.9 (5.5)</td>
<td>0.094</td>
</tr>
<tr>
<td>BDS 5 All non-participating</td>
<td>146</td>
<td>65.5 (5.9)</td>
<td></td>
</tr>
</tbody>
</table>

Selected students’ peer-DOPS feedback comments and the subsequent agreed action plans for both study groups are presented in Table 3-10. Further, the majority of BDS 2 and BDS 5 students expressed a positive perception of the prospective peer-assessment and peer-feedback protocol as presented in Table 3-11.
Table 3-10 Selected BDS 2 and BDS 5 students’ feedback comments and agreed challenges and actions, extracted from their peer-DOPS completed forms.

<table>
<thead>
<tr>
<th>Feedback comments from peers</th>
<th>Agreed challenges and actions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BDS 2</strong></td>
<td></td>
</tr>
<tr>
<td>Good work, maybe more packing could be used with amalgam</td>
<td>Better condensation of amalgam</td>
</tr>
<tr>
<td>Shape and anatomy of amalgam must be improved</td>
<td>Better / improved fissure pattern</td>
</tr>
<tr>
<td>Need specific targeted advice for improving positioning for vision in phantom head – tending to working outside head, freehand.</td>
<td>Work more together with demonstrator to facilitate affective practice in the phantom head</td>
</tr>
<tr>
<td>Difficulty with indirect vision e.g. Direct for posteriors</td>
<td>Practice using dental mirrors to get used to indirect vision</td>
</tr>
<tr>
<td>Margin of restoration could have been smoother</td>
<td>Cavity margins smoother and better anatomy on next restoration</td>
</tr>
<tr>
<td>Time management lacking</td>
<td>To be more efficient, aim to two teeth per session</td>
</tr>
<tr>
<td><strong>BDS 5</strong></td>
<td></td>
</tr>
<tr>
<td>Very difficult case. Could be better listening to patient and own time management</td>
<td>In the future try not to undertake work beyond own capabilities and ensure complexity of dentistry is understood</td>
</tr>
<tr>
<td>Being more confident in own ability</td>
<td>Correct laboratory instructions</td>
</tr>
<tr>
<td>Trouble placing temporary crown on with Temp Bond</td>
<td>Better control throughout</td>
</tr>
<tr>
<td>Problem placing rubber dam. Place the hole more centrally for full coverage</td>
<td>Improve technique</td>
</tr>
<tr>
<td>Could improve infection control procedures, technically and clinically</td>
<td>Better cross infection control</td>
</tr>
<tr>
<td>Student a little out of practice in jaw registration of partial dentures</td>
<td>Read up on jaw registration of partials</td>
</tr>
</tbody>
</table>

Table 3-11 Anonymous BDS 2 (n=10) and BDS 5 (n=16) students’ perceptions (%) of the peer-assessment protocol after using the peer-DOPS instruments for 6 and 7 occasion, respectively.

<table>
<thead>
<tr>
<th>To what extent do you agree that the peer-assessment and feedback protocol used in this study:</th>
<th>Course</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Nor Agree nor disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>...assessed you in areas that correspond to your activity in the pre-clinic/clinic?</td>
<td>BDS 2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>70.0</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>37.5</td>
<td>62.5</td>
</tr>
<tr>
<td>...could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?</td>
<td>BDS 2</td>
<td>0</td>
<td>0</td>
<td>20.0</td>
<td>40.0</td>
<td>40.0</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0</td>
<td>0</td>
<td>12.5</td>
<td>12.5</td>
<td>75</td>
</tr>
<tr>
<td>...have helped you to identify learning needs and to improve your performance?</td>
<td>BDS 2</td>
<td>0</td>
<td>0</td>
<td>30.0</td>
<td>50.0</td>
<td>20.0</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0</td>
<td>0</td>
<td>12.5</td>
<td>50.0</td>
<td>37.5</td>
</tr>
<tr>
<td>...was acceptable and fair?</td>
<td>BDS 2</td>
<td>0</td>
<td>0</td>
<td>20.0</td>
<td>50.0</td>
<td>30.0</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18.8</td>
<td>81.3</td>
</tr>
</tbody>
</table>
3.6 Discussion

The present Chapter reports the development and piloting of a structured protocol of formative prospective peer-assessment of undergraduate pre-clinical and clinical dental students’ skills, used as an informed framework for the provision of immediate peer-feedback using two newly generated peer-assessment tools.

Accordingly, BDS 2 and BDS 5 students used direct observation as a method of collecting information, guided by their peer-assessment tools as micro-frameworks (Pangaro and ten Cate, 2013), to make judgements of their peers’ process of simulation exercises (pre-clinical students) and patient care (clinical students). In contrast to what staff trainers do, they did not judge their peer’s competence or performance “quality” nor the practice “volume” (Norcini, 2003a).

Both BDS 2 and BDS 5 overall mean peer-DOPS scores (4.8 and 5.6, respectively, Table 3-6) were higher than the reported 4.55 (sd=0.48) by Archer et al. (2008) when 553 medical Foundation trainees were peer-assessed once using the Mini-PAT (25% were assessed twice). Despite these differences in means, peer-scores ranged from similar minimum values of 2 and 2.13, to maximum ones of 6 and 5.77, for our and their study, respectively. However, this was not the case in the study by Bennett et al. (2012), where 40 undergraduate medical students assessed each other’s diagnostics and/or treatment plan skills using a standard Mini-CEX form and reported no students’ scores lower than 4 out of 5, which they stated was unlikely to accurately reflect the range of their students’ performance.

The reliability coefficient G was higher for BDS 5 clinical peer-DOPS (0.67 for seven occasions) as compared to BDS 2 pre-clinical peer-DOPS (0.62 for six occasions). These G coefficients are comparable to those reported by Wilkinson et al. (2008) when six Multi-Source Feedback (MSF) encounters (0.65) were used to assess medical specialists. However, they are lower than those reported by Moonen-van Loon et al. (2013) for six (0.74) and seven (0.76) MSF residents’ encounters performed by supervisors, peers, nurses, administrative staff, patients, and self-assessment. Further, they are also lower than the generally accepted G coefficient of ≥0.7 as sufficient for ‘low-stakes’ assessment situations (Beard et al., 2011). A possible explanation for this might be found in the number of peer-assessment encounters completed by the students. BDS 2 only reached a mean of 5.7 encounters per students, while BDS 5 reached a mean of 6.5 per students. Both these cases are lower than the minimum of 7 to 11
judgements required for reliable findings as concluded by Williams et al. (2003). A longer and larger peer-assessment study using the current protocol might be needed to evaluate the effect of more students’ encounters on the reliability of the instruments.

Despite both groups of undergraduate students being on the path from novice to experts (Dreyfus et al., 1988), and against the Sluijsmans et al. (2002) hypothesis that novices are less capable of assessing certain domains, the analysis of the current data showed that both groups were able to identify differences in competence and performance when assessing their peers across the respective eleven and thirteen domain peer-DOPS forms. This being the case, the first hypothesis of this Chapter can be accepted: The developed prospective peer-assessment protocol reliably allows pre-clinical dental students to identify competence differences in various domains. Similarly, the second hypothesis can also be accepted: The developed prospective peer-assessment protocol reliably allows clinical dental students to identify performance differences in various domains.

Accordingly, BDS 2 students distinguished between a better clinical skill in “Observing aseptic technique...” and a lower professional behaviour in “Does the trainee show insight into his/her performance?” (Table 3-7). Whilst the former (mean 5.3) corresponds well with the intensive training students have in infection control and safety starting as early as BDS 1, the latter (mean 4.5) was not expected. Thus, despite the natural motivation for “friendship” marking as described by Norcini (2003b), peers scored each other significantly low (p=0.04) in this professionalism item. In contrast to a possible blandness, this might be better explained by a competitive spirit as illustrated by Cushing et al. (2011), or as a real sign that these junior students are still developing this self-assessment skills. This BDS 2 item low scoring was not observed in the same item of senior BDS 5 students’ peer-assessment. By way of contrast, BDS 5 students assessed the professional behaviour of “Consideration of patient/ professionalism” item with the maximum possible score (mean 6), demonstrating a high social commitment. At the same time, several other items in the clinical (items 1 & 6), management (item 8) communication (item 9), domains were significantly lower. These findings are in agreement with the study of Bennett et al. (2012), who reported medical undergraduates’ peer-assessment ability to identify “areas in which peers performed well and those that required improvement”.

Yet another indication of the students’ ability to make an accurate inference of their
peer’ performance (Streiner and Norman, 2008) was their ability to detect changes in their peers’ performance with time, though with dissimilar patterns (Figure 3-7). In the case of BDS 5 students they were able to perceive significant progress in their peers’ performance by the third peer-assessment session. In contrast, BDS 2 students only started to notice peer improvement at the fifth session, while during the first four encounters scores dropped significantly. This difference might be explained by an initial calibration process (Hauser and Bowen, 2009), or the time needed to gain experience as evaluators (Karl et al., 2011), or even adjusting to the learning environment (Schoenrock-Adema et al., 2007), before scores started to rise. This increase in ratings over time is in keeping with the earlier findings of Prescott-Clements et al. (2011) in postgraduate dental trainees assessed by staff members in eight similar items, and Davies et al. (2009) in foundation medical trainees’ peer-assessment.

Further, whilst not statistically significant, both groups’ peer-assessment scores were positively correlated to students’ end-of-year examination marks. Once again, a larger peer-assessment study of the current protocol might be needed to better evaluate this validity criterion.

Similarly, despite students’ high perceptions in that the peer-DOPS protocol “helped them to identify learning needs and improve performance” (Table 3-11), it had no observable educational impact as both groups of participating students show no difference in high stakes end-of-year examination marks when compared to the other students of their respective classes. This might be explained once again by the short observation period (around three months). Already graduated foundation medical trainees’ peer-assessment (using the mini-PAT), required two consecutive semester assessments to detect a slight increase in scores from 4.6 to 4.7 (scale from 1 to 6) (Davies et al., 2009).

At the end of the piloting phase of the study, an anonymous questionnaire of both BDS 2 and BDS 5 groups of students gathered their perception of the peer-assessment protocol (Table 3-11). Most of them expressed their positive perceptions of the peer-assessment and peer-feedback protocol, which allows for the third hypothesis of this Chapter to be accepted: Most students have positive perceptions of the peer-assessment and peer-feedback protocol.

Similar positive perceptions were recently reported by Kirton et al. (2013) when 90.8%
of 643 dental Foundation trainees, assessed by trainers, found a longitudinal evaluation of performance (D-EPs), useful.

All students (100%) agreed or strongly agreed that both peer-DOPS instruments assessed suitable domains. Likewise, the vast majority (>80%) agreed or strongly agreed that the peer-DOPS protocol was feasible to be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education; helped them to identify learning needs and improve performance; and probably most importantly, was acceptable and fair. The latter is particularly important as a sound assessment needs to be accepted, fair (Norman et al., 1991) and authentic (McCoubrie, 2004) or it is destined to fail (van der Vleuten, 1996). The fact that students’ views and opinions were requested and their answers implemented during the development phase of the peer-assessment instruments, may have contributed to an increase in their acceptability (Shue et al., 2005). Further, we did not observe nor receive any negative comment or concern related to the peer-assessment experience like students not accepting peer feedback as accurate or regarding it as of little benefit, inhibited or constrained students, inappropriate and hurtful comments or any other sign of peer power relationships, as previously reported (Van Rosendaal and Jennett, 1992; Topping, 1998; Dannefer et al., 2005; Boud and Falchikov, 2007).

The present pilot results are highly significant and interesting inasmuch as there is no data in the literature on the prospective use of peer-assessment in the dental undergraduate setting. However, the development and piloting of the protocol was very time-consuming and required careful and rigorous preparation as well as dedication. Further, the limiting sample size and the short observation period, means care should be exercised when trying to generalise.

3.7 Conclusion

This Chapter developing and piloting of a prospective peer-assessment protocol has established the ability of undergraduate dental students to reliably and accurately peer-assess one another in all four GDC domains. Further, it has shown that the approach was positively perceived by participating students.

However, a further implementation of the piloted peer-assessment tools over a longer period of time and across a wider study body is necessary to provide more reliable and
generalizable data. Thus, questions mainly related to the tools’ utility as an assessment method (van der Vleuten, 1996), and the educational effects of the protocol’s combined peer-assessment and peer-feedback, followed by self-assessment and reflection, remain unanswered. Consequently, the following chapter will describe the usefulness of a method to assess students’ reflection skills in order to later, in the subsequent peer-assessment implementation phase (Chapter 5), evaluate the potential of the protocol to foster students’ higher order thinking skills.
Chapter 4

Utility of a Method to Assess Dental Students’ Reflection Habits

Figure 4-1 Flowchart of the study’s research chapters. The fourth chapter is highlighted to help in providing an overall view.

4.1 Introduction

Reflection, as discussed in Chapter 1 (Reflection and learning in page 77), has long been acknowledged as an important concept in the learning process (Dewey, 1938; Boud et al., 1985) and as on-going activity, it is regarded as a crucial and essential characteristic of healthcare professionals’ education and competent professional practice (Mann et al., 2009; Aronson, 2011; Chambers et al., 2011).

Thus, in an educational environment, it would be useful to determine whether students
or trainees are engaged in reflective practice. This requires some means of identifying reflective thought and a measure of its depth (Kember, 1999; Pee et al., 2002). However, there appears to be no current reflective practice assessment procedure with sufficient validity, acceptability or clear formulation which would qualify it as the gold standard for assessing students’ reflective practice abilities (National Postsecondary Education Cooperative, 2000; Chambers et al., 2011).

Consequently, several approaches have been proposed to measure students’ reflectivity (Pee et al., 2002). Thus, the majority of these reported methods in health sciences consist of various forms of reflective writings in medicine (Niemi, 1997; Driessen et al., 2005; Learman et al., 2008; Wald et al., 2009; Aronson et al., 2012; Dekker et al., 2013); dentistry (Pee et al., 2002; Boyd, 2008; Bush and Bissell, 2008; Tsang and Walsh, 2010; Jonas-Dwyer et al., 2013; Tonni and Oliver, 2013a; Wallace et al., 2013); nursing (Wong et al., 1995); and physiotherapy (Williams et al., 2002). This despite reports that students dislike “writing” reflections (Bush and Bissell, 2008), and that their academic pressure discourage them from engaging in honest and open reflection (Hargreaves, 2004),

The literature also includes accounts of self-reported questionnaires/surveys (Mitchell, 1994; Kember et al., 2000; Sobral, 2000; Aukes et al., 2007), reflective blogging (Wetmore et al., 2010), semi-structured questionnaires with reflection-evoking case vignettes (Boenink et al., 2004), surveys of attitudes and the use of reflective portfolios (Pearson and Heywood, 2004; Bush and Bissell, 2008; Kardos et al., 2009; Koole et al., 2013), and interviews (Hallett, 1997; Boyd, 2008).

The vast majority of these methods were designed and used as reflective learning activities in an attempt to foster students’ reflective thinking, whilst the assessment and depth of the reflective thoughts followed a demanding qualitative approach. However, for the purpose of the present chapter study, a feasible method of assessing reflection skills that could be implemented in large numbers of students and trainees, both before and after the peer-assessment implementation, was required.

Further, it should also allow for comparisons of pre- and post-intervention reflection skills without influencing the sample’s reflective thoughts, as well as permitting statistical correlation with other interval scales. In other words, we needed a viable, quantitative and diagnostic approach to assess reflection skills.
In view of this, we considered using the self-reported questionnaire from Sobral (2000), the Gröningen Reflection Ability Scale (GRAS) (Aukes et al., 2007), and the self-reported reflection questionnaire from Kember et al. (2000). However, keeping in view its confirmed psychometric properties (Lethbridge et al., 2011) and its extensive use in multiple disciplines (Phan, 2009; Lim, 2011; Chelliah and Arumugam, 2012; Naghdipour and Emeagwali, 2013; Ambrose and Ker, 2014), we decided that the latter was most appropriate.

4.2 Aim

The aim of this Chapter was to assess the usefulness, in terms of reliability and construct validity of the self-reported Reflection Questionnaire (RQ) (Kember et al., 2000) as a method of assessing dental students’ and postgraduate trainees’ reflection habits.

4.3 Hypotheses

i. The results of the studied RQ are reliable.

ii. The studied RQ scores discriminate between subjects’ non-reflective and reflective thoughts.

iii. The levels of reflective habits from undergraduate students and postgraduate trainees are the same.

iv. Undergraduate dental students’ reflective habits and their high stakes academic achievements are positively correlated.

4.4 Materials and Methods

4.4.1 The instrument

The RQ of Kember et al. (2000) was developed as a simple diagnostic instrument to examine students’ engagement in reflective thinking in professional academic programmes from any discipline. Thus, the questionnaire, based on Mezirow’s (1991) conceptualisation of levels of reflective thinking, has 16-item (statements) that conform
four scales or factors to quantitatively assesses two levels of non-reflective actions (Habitual Action and Understanding), and two levels of reflective actions (Reflection and Critical Reflection).

According to Kember et al. (2008) the descriptions for each of the four categories are:

- **Habitual Action**: previously learnt responses that are automatically performed with little conscious thought. It is what expert practitioners do in routine practice as they have dealt with the same situation many times before. It is also what novice learners do when they rigidly follow the steps of taught procedures without thinking its applicability or alternatives. Habitual Action also occurs when students provide an answer without trying to understand or forming a view of the concept, task or theory that underpins the topic, which is compatible with a surface approach to learning.

- **Understanding**: a cognitive learning and/or reading but without appraising the concept or task. Still, the student attempts to reach an understanding and meaning of the theory, and therefore is related to a deep approach to learning. However, the concepts are not related to personal experiences or real applications and, as such, do not have personal meaning, may not be incorporated as new knowledge, and its retention can be limited. Understanding is usually a characteristic of inexperienced undergraduate students thinking, as they lack practice to apply the new knowledge.

- **Reflection**: an active, persistent, and careful critique of assumptions about the content or process of problem solving. It is the next level after understanding as the concepts or tasks are related to previous knowledge and personal experiences. Theory then is interpreted and practically applied.

- **Critical Reflection**: becoming aware of why we perceive, think, feel or act as we do. It implies undergoing a transformation of perspective which requires first a critical review of presuppositions and assumptions from prior learning and their consequences. These are not easy to change and therefore, critical reflection is unlikely to happen regularly, especially in professionals. However, it is expected to be more common in students who are learning a subject or profession and so are establishing their assumptions and constructing their individual meanings.

When answering the RQ, participants are asked to indicate their level of agreement with all 16-items on a five-point Likert scale ranging from “definitely disagree” (score as 1), “disagree with reservations” (score as 2), “neutral” (score as 3), “agree with
reservations” (score as 4), and “definitely agree” (score as 5) (Table 4-1).

Table 4-1 Reflection Questionnaire from Kember et al. (2000) used to assess the reflective habits of students and trainees.

<table>
<thead>
<tr>
<th>Statements</th>
<th>Definitely agree</th>
<th>Agree with reservation</th>
<th>Neutral</th>
<th>Disagree with reservation</th>
<th>Definitely disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. When I am working on some activities, I can do them without thinking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Dental training requires me to understand concepts taught by the</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. I sometimes question the way others do something and try to think of</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. As a result of my dental training I have changed the way I look at</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. In dental training I do things so many times that I have started</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. To pass dental training examinations you need to understand the content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. I like to think over what I have been doing and consider alternative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Material learned during my dental training has challenged some of my</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. As long as I can remember material presented in class for examinations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. I need to understand the material taught by the teacher in order to</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. I often reflect on my actions to see whether I could have improved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. As a result of the material learned in dental training I have</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. If I follow what the lecturers say, I do not have to think too much</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. In dental training you have to continually think about the material</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. I often re-appraise my experience so I can learn from it and improve</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. During dental training I discovered faults in what I had previously</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Each of the four scales is measured by four contributory items, and the participant’s score on each of these dimensions is calculated by adding the responses score for each of the four items from each scale. Once completed, each participant receives a score ranging from a minimum of 4 (definitely disagree in all four items of that scale) to a
maximum of 20 (definitely agree in all four items of that scale) for each Habitual Action (statements 1+5+9+13), Understanding (statements 2+6+10+14), Reflection (statements 3+7+11+15) and Critical Reflection (statements 4+8+12+16) dimensions (Table 4-1). There is no overall score as this would not be consistent with the theoretical underpinnings of the scale development (Kember et al., 2000). Consequently, the higher the score in each category, the more agreement with engaging in the particular dimension that each scale measures. Lastly, the RQ, along with an Information Sheet explaining the study, a Consent Form as well as a final free text box for comments, were developed electronically and the RQ’s optimal online technical performance was tested using Survey Monkey®.

4.4.2 Participants and data collection

At the end of June 2012, all 740 BDS years 1 to 5 undergraduate students and 221 postgraduate clinical dentistry programme and PhD postgraduate trainees enrolled at KCLDI, received an electronic invitation from the Director of Education or Director of Graduate Studies at KCLDI to voluntarily participate in the study by completing the online RQ.

4.4.3 Data analysis

Participation was closed twelve weeks after the invitation was sent. Completed questionnaire responses were downloaded from Survey Monkey® into a spreadsheet and imported into SPSS® version 21 for analysis.

Descriptive statistics were employed to express participants’ gender and age characteristics as well as their current and previous academic backgrounds. Normal distribution of the resultant RQ scores was assessed by visual inspection of a histogram and by exploring the central tendency (mean, median, and mode). The internal reliability of the items constituting each scale (Habitual Action, Understanding, Reflection and Critical Reflection) was quantified by computing Cronbach’s alpha coefficient.

Construct validity of the RQ was assessed in five different ways:

- To test whether the questionnaire could distinguish between participants’ four scales of reflective skills, a one-way analysis of variance (ANOVA) with post hoc analysis using Tukey’s test, was performed.
• To assess the RQ capability to differentiate between participants’ four scales of reflective skills according to their academic course (BDS 1 to 5, clinical postgraduate, and PhD), a one-way analysis of variance (ANOVA) with post hoc analysis using Tukey’s test, was performed.

• Subjects’ reflection skills for each of the RQ four scales were compared according to their gender, age group (≤23 years old and ≥24 years old) and whether they had a previous university degree, using separate independent samples t-tests.

• Inter-scale relationships between all four RQ scales, as predicted conceptually by the questionnaire developers (Kember et al., 2000), were studied using a Pearson correlation analysis.

• To establish the relationship between the reflection habits of undergraduate subjects and their academic achievements a Pearson correlation analysis was performed. This was measured by collecting students’ end-of-year high stakes examination marks and subsequently correlating them to their RQ scores.

To enable the fifth analysis, examination marks from Paper 1a for BDS 1 to 3 (short answer questions and short note questions), Paper 1b for BDS 4 and 5 (written essays on clinical scenarios), Paper 2 (online MCQ), an OSCE, Clinical Reasoning Examination (CRE), and Case Presentation were obtained and correlated to students’ Habitual Action, Understanding, Reflection and Critical Reflection scores. It should be mentioned that all students take Papers 1 and 2 whereas the OSCE is used to assess BDS Years 3 to 5, and CRE along with Case Presentation are part of BDS Year 5 Finals.

4.5 Results

4.5.1 Participants and data collection

A total of 324 (34%) undergraduate students and post-graduate trainees (208 females and 116 males) answered the invitation and completed all 16 items of the online self-reported RQ. No issues were reported by any student or trainee and only one exited the questionnaire before completing all 16 statements. However, the same student later fully completed it in the second attempt.
Twenty-four undergraduate students submitted a free text comment at the end of the questionnaire. Two hundred and eighty-one participants were undergraduate students representing all BDS courses (years 1 to 5), while 43 were postgraduate trainees following a clinical dentistry programme (N=25) or a PhD (N=18). At the time the study was conducted, all groups of respondents had an age range between 18 and 46 years (mean=25, sd=6.7), while 44% of them were 23 years of age or younger. The number of participants who had a previous national or international university degree was 131, representing 40% of the whole sample (Table 4-2).

<table>
<thead>
<tr>
<th>Cohort</th>
<th>N</th>
<th>Mean age (sd)</th>
<th>Previous degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS 1</td>
<td>38</td>
<td>20.5 (2.5)</td>
<td>10%</td>
</tr>
<tr>
<td>BDS 2</td>
<td>71</td>
<td>22.7 (4.8)</td>
<td>25%</td>
</tr>
<tr>
<td>BDS 3</td>
<td>44</td>
<td>24.4 (5.9)</td>
<td>53%</td>
</tr>
<tr>
<td>BDS 4</td>
<td>42</td>
<td>23.9 (3.0)</td>
<td>23%</td>
</tr>
<tr>
<td>BDS 5</td>
<td>86</td>
<td>24.9 (2.8)</td>
<td>39%</td>
</tr>
<tr>
<td>Clinical Postgraduate (1st, 2nd and 3rd year)</td>
<td>25</td>
<td>36.5 (5.7)</td>
<td>100%</td>
</tr>
<tr>
<td>PhD (2nd and 3rd year)</td>
<td>18</td>
<td>32.4 (7.1)</td>
<td>100%</td>
</tr>
<tr>
<td>Totals</td>
<td>324</td>
<td>25.0 (6.7)</td>
<td>40%</td>
</tr>
</tbody>
</table>

**4.5.2 Data analysis**

Reflection questionnaire answers ranged from 1 (definitely disagree) to 5 (definitely agree) (mean=3.7, sd=0.9, median=4.0, mode=4) and were normally distributed. Cronbach’s alpha internal reliability coefficient was 0.696 for Habitual Action, 0.711 for Understanding, 0.702 for Reflection and 0.712 for Critical Reflection.

Resultant RQ four-scale mean scores from all participants was 10.7 (sd=1.6) for Habitual Action, 17.3 (sd=1.5) for Understanding, 17.0 (sd=1.8) for Reflection, and 14.4 (sd=2.0) for Critical Reflection. The ANOVA test confirmed these scores as significantly different ($F=845.02$, $p<0.0001$). Hence, a further post hoc analysis using Tukey’s test was undertaken (Table 4-3).

Interestingly, 5.8% of undergraduate students exhibited a Reflection score under two standard deviations (13.4) of the overall mean (17.0). The percentage of students lower than two standard deviations (10.2) of the overall mean for Critical Reflection (14.4) increased to 8.0%. However, no postgraduate or PhD trainees were found in these groups.
Table 4-3 Mean value for each of the four RQ scales of all subjects and ANOVA with post hoc Tukey’s test results showing inter-scale mean difference and corresponding p-values.

<table>
<thead>
<tr>
<th>Reflection Questionnaire scales</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale Mean</td>
<td>10.7</td>
<td>17.3</td>
<td>17.0</td>
<td>14.4</td>
</tr>
<tr>
<td>Mean difference</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>p value of diff.</td>
<td>&lt;0.0001</td>
<td>0.302</td>
<td>0.148</td>
<td>-</td>
</tr>
<tr>
<td>Habitantial Action</td>
<td>10.7</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Understanding</td>
<td>17.3</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Reflection</td>
<td>17.0</td>
<td>0.302</td>
<td>0.148</td>
<td>-</td>
</tr>
<tr>
<td>Critical Reflection</td>
<td>14.4</td>
<td>2.980</td>
<td>2.679</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

The mean scores for each of the RQ scales according to participants’ academic course and study year were calculated (Table 4-4), and graphically illustrates (Figure 4-2), grouping all undergraduate students together to compare them with clinical postgraduates and PhD trainees. The ANOVA test found no significant differences between these three groups in Habitual Action ($F=0.82$, $p=0.921$), and Understanding ($F=2.760$, $p=0.067$).

Table 4-4 Mean scores and standard deviations (sd) for each of the four scales of the RQ according to participants’ academic course and study year.

<table>
<thead>
<tr>
<th>Reflection Questionnaire scales</th>
<th>Cohort</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td>BDS 1</td>
<td>38</td>
<td>9.2 (2.1)</td>
<td>17.2 (1.7)</td>
<td>16.2 (2.0)</td>
<td>12.7 (2.1)</td>
</tr>
<tr>
<td>BDS 2</td>
<td>71</td>
<td>9.9 (2.3)</td>
<td>17.8 (1.2)</td>
<td>16.4 (2.1)</td>
<td>13.1 (2.3)</td>
</tr>
<tr>
<td>BDS 3</td>
<td>44</td>
<td>10.7 (1.7)</td>
<td>16.9 (1.6)</td>
<td>16.5 (1.7)</td>
<td>13.6 (1.9)</td>
</tr>
<tr>
<td>BDS 4</td>
<td>42</td>
<td>10.5 (1.5)</td>
<td>17.0 (1.7)</td>
<td>17.1 (1.9)</td>
<td>14.8 (1.8)</td>
</tr>
<tr>
<td>BDS 5</td>
<td>86</td>
<td>12.4 (1.4)</td>
<td>17.2 (2.2)</td>
<td>17.5 (1.3)</td>
<td>15.5 (1.7)</td>
</tr>
<tr>
<td>Clinical Postgraduate</td>
<td>25</td>
<td>10.3 (1.2)</td>
<td>17.5 (1.1)</td>
<td>17.6 (1.2)</td>
<td>16.3 (1.2)</td>
</tr>
<tr>
<td>PhD</td>
<td>18</td>
<td>10.4 (1.0)</td>
<td>18.5 (1.6)</td>
<td>18.3 (1.5)</td>
<td>18.1 (1.1)</td>
</tr>
<tr>
<td>Scale Mean</td>
<td>324</td>
<td>10.7 (1.6)</td>
<td>17.3 (1.5)</td>
<td>17.0 (1.8)</td>
<td>14.4 (2.0)</td>
</tr>
</tbody>
</table>

Conversely, the same test indicated significant differences between the groups in Reflection ($F=3.725$, $p=0.027$), with the post hoc Tukey’s test showing undergraduate
student’s (mean 17.0) and PhD trainee’s (mean 18.3) scores differed significantly ($p=0.038$). However, the most significant difference was seen in the Critical Reflection scale ($F=15.312$, $p<0.0001$), where the post hoc Tukey’s test revealed the differences between undergraduate student’s scores (mean 14.1) and clinical postgraduate trainee’s ones (mean 16.3) were significant ($p=0.036$), as well as that from undergraduates’ and PhD trainees’ (mean 18.1) ($p<0.0001$).

Comparisons between Habitual Action, Understanding, Reflection and Critical Reflection mean scores of the sampled subjects grouped by gender, age group and whether they had a previous university degree, found statistically significant differences (Table 4-5). Accordingly, the ≥ 24 years old group showed higher scores than those ≤ 23 years old both in Reflection (17.5 versus 16.4, $p<0.0003$) and Critical Reflection (15.1 versus 13.6, $p<0.0006$) scales. Similarly, those who attained a previous university degree exhibited higher scores those who did not in Reflection (17.5 versus 16.7, $p<0.011$) and Critical Reflection (15.3 versus 13.8, $p<0.0004$).

![Figure 4-2](image-url)
Chapter 4 Utility of a Method to Assess Dental Students’ Reflection Habits

Table 4-5 Separate independent samples t-test results showing number, mean scores, standard deviations (sd) and p-values for each of the four RQ scales according to participants’ gender, age group and whether they had a previous university degree.

<table>
<thead>
<tr>
<th>Reflection Questionnaire scales</th>
<th>Cohort</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean (sd)</td>
<td>p value of diff.</td>
<td>Mean (sd)</td>
<td>p value of diff.</td>
</tr>
<tr>
<td>Female</td>
<td>208</td>
<td>10.7 (2.9)</td>
<td>0.948</td>
<td>17.4 (2.1)</td>
<td>0.476</td>
</tr>
<tr>
<td>Male</td>
<td>116</td>
<td>10.7 (2.6)</td>
<td>17.2 (2.0)</td>
<td>0.711</td>
<td>17.2 (2.0)</td>
</tr>
<tr>
<td>≤ 23 years old</td>
<td>143</td>
<td>10.8 (2.6)</td>
<td>17.4 (2.1)</td>
<td>0.090</td>
<td>17.6 (1.8)</td>
</tr>
<tr>
<td>≥ 24 years old</td>
<td>181</td>
<td>10.6 (2.9)</td>
<td>17.1 (2.1)</td>
<td>0.065</td>
<td>17.6 (2.1)</td>
</tr>
<tr>
<td>Previous degree</td>
<td>131</td>
<td>10.4 (2.7)</td>
<td>17.0 (2.2)</td>
<td>0.046</td>
<td>17.0 (2.2)</td>
</tr>
<tr>
<td>No previous degree</td>
<td>193</td>
<td>10.9 (2.8)</td>
<td>17.2 (2.2)</td>
<td>0.044</td>
<td>17.2 (2.2)</td>
</tr>
</tbody>
</table>

The tabulation of the RQ inter-scale correlations and their level of statistical significance are shown in Table 4-6 below. Understanding and Reflection showed the highest relationships, while Habitual Action and Critical Reflection exhibited the lowest.

Table 4-6 Mean value for each of the four RQ scales of all subjects and inter-scale Pearson correlations with the corresponding p-value.

<table>
<thead>
<tr>
<th>Reflection Questionnaire scales</th>
<th>Mean</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitual Action</td>
<td>10.7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Understanding</td>
<td>17.3</td>
<td>0.046</td>
<td>0.63</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Reflection</td>
<td>17.0</td>
<td>0.423</td>
<td>&lt;0.0001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Critical Reflection</td>
<td>14.4</td>
<td>0.354</td>
<td>0.003</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Correlation analysis between reflective habits of undergraduate subjects and their academic performances in all end-of-year high stakes examinations showed wide diversity of relationships between RQ scales and the different assessment programme methods and is particularly interesting in these results (Table 4-7).
Table 4-7 Correlation and corresponding p-value between each RQ scale of BDS year 1-2-3-4-5 subjects and their academic performances in end-of-year examinations. For each examination, the cohort, number of subjects and the examination mean score of the sample as well as that from the whole class ( ) are shown.

<table>
<thead>
<tr>
<th>Examination</th>
<th>Reflection Questionnaire scales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Habitual Action</td>
</tr>
<tr>
<td></td>
<td>Correlation</td>
</tr>
<tr>
<td>Paper 1a (short answer questions, short note questions)</td>
<td>BDS 1 to 3</td>
</tr>
<tr>
<td>Paper 1b (written essays on clinical scenarios)</td>
<td>BDS 4 &amp; 5</td>
</tr>
<tr>
<td>Paper 2 (online multiple choice questions)</td>
<td>BDS 1 to 5</td>
</tr>
<tr>
<td>OSCE</td>
<td>BDS 3 to 5</td>
</tr>
<tr>
<td>Clinical Reasoning</td>
<td>BDS 5</td>
</tr>
<tr>
<td>Case Presentation</td>
<td>BDS 5</td>
</tr>
</tbody>
</table>

Selected free text comments one from each BDS course that were submitted by participants after completing the RQ are displayed in Table 4-8, as a qualitative complement of the data presented above.

Table 4-8 Selection of participants’ comments, study cohort, age, previous degree and their scores in Habitual Action (HA), Understanding (U), Reflection (R), and Critical Reflection (CR).

<table>
<thead>
<tr>
<th>Comment</th>
<th>Study group</th>
<th>Age</th>
<th>Previous degree</th>
<th>HA</th>
<th>U</th>
<th>R</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>“From the questions asked in this questionnaire, I will try and reflect more on my actions in the clinic so that I can improve the next time”</td>
<td>BDS 1</td>
<td>21</td>
<td>No</td>
<td>6</td>
<td>18</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>“Sometimes when it comes to methods, instead of thinking of better ways to do things I think to what is expected of me in exams”</td>
<td>BDS 2</td>
<td>22</td>
<td>No</td>
<td>12</td>
<td>19</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>“I keep a dentistry diary where I record what I have learned &amp; what I can improve on in my clinical sessions and all the small bits advice I get from my tutors which are specific to the treatment situation that would not have been mentioned in the main lectures - I find this really helps build my knowledge and confidence for the next session - and also looking back on the past entries I can see how far I have already progressed as a student dentist”</td>
<td>BDS 3</td>
<td>35</td>
<td>Yes</td>
<td>5</td>
<td>20</td>
<td>19</td>
<td>19</td>
</tr>
</tbody>
</table>
4.6 Discussion

The present Chapter describes the usefulness of the self-reported Kember’s Reflection Questionnaire as a method to assess dental students’ and trainees’ reflection habits. Taking into consideration that peer-assessment can encourage self-reflection (Schön, 1983; Boud and Falchikov, 2007), and that reflection is key to learning (Boud et al., 1985), the formative peer-assessment and peer-feedback protocol needs to identify whether the peers’ exercise is helping students to engage in reflective practice and, ideally, its depth. The results of the present Chapter show the various levels of self-reported reflective thinking habits of undergraduate dental students and postgraduate trainees following clinical and PhD postgraduate programmes. Further, it details the relationship between these reflective habits and the academic performance of undergraduate students.

The internal reliability coefficients of the RQ scores were acceptable and consistent with previous reports of the same instruments used in students from health sciences (Occupational therapy, Physiotherapy, Radiography and Nursing) (Kember et al., 2000; Leung and Kember, 2003), nursing (Lethbridge et al., 2011), polytechnic (Lim, 2011), medicine (Ambrose and Ker, 2014), and arts and mathematics (Phan, 2008). These results allow for the first hypothesis of this Chapter to be accepted: The results of the studied RQ are reliable.

The mean of all subjects’ scores showed the RQ did differentiate between non-reflective and reflective thoughts. Thus, the non-reflective skill of Habitual Action (10.7) was significantly different ($p<0.0001$) with both reflective skills of Reflection (17.0) and Critical Reflection (14.4). Similarly, the non-reflective thought of Understanding (17.3)
was significantly different \((p<0.0001)\) with the reflective ones of Critical Reflection \((14.4)\) \((\text{Table 4-3})\). Consequently, this discernment permits to accept the second hypothesis of the present Chapter: The studied RQ scores discriminate between subjects’ non-reflective and reflective thoughts.

The higher Understanding \((17.3)\) and Reflection \((17.0)\) mean scores indicate that students and postgraduate trainees employed these levels of thinking more than Habitual Action and Critical Reflection \((\text{Kember et al., 2000; Lethbridge et al., 2011; Chelliah and Arumugam, 2012})\). Indeed, Ashley \textit{et al.} (2006) reported that BDS Year 4 students placed a “great deal of emphasis on practical applications of their knowledge”. Thus, consistent with a university education, they would predominantly be learning – Understanding – and integrating this new knowledge into their practice – Reflecting. On the other hand, not only would Habitual Action, characterised by its automated responses, be uncommon in busy and active learning courses, but also Critical Reflection, as it requires a major change of perspective and alteration to deep-seated beliefs \((\text{Kember et al., 2000})\).

Reflection and Critical Reflection scores increased more than those from Habitual Action and Understanding as students progressed through the programme. As previously reported \((\text{Chelliah and Arumugam, 2012; Ambrose and Ker, 2014})\), this might indicate that theory and concept integration into students’ and trainees’ practice increases as they climb the programme ladder. Further, it is acknowledged that dental students are required to develop reflective thinking and problem solving skills as they move from well-defined problems in the classroom environment to the more uncertain and ill-defined real-life situations when they start clinical patient care \((\text{Boyd, 2008})\).

The undergraduate students’ mean scores for Understanding, Reflection and Critical Reflection were higher, while Habitual Action ones were lower than those reported for Hong Kong undergraduate students in occupational therapy, physiotherapy, radiography and nursing \((\text{Kember et al., 2000})\), Hong Kong health science students \((\text{Leung and Kember, 2003})\), Malaysian and Chinese year 3 and year 4 medical students \((\text{Chelliah and Arumugam, 2012})\), and those studying English language teaching in Northern Cyprus \((\text{Naghdipour and Emeagwali, 2013})\). Indeed, our undergraduate RQ mean scores were more comparable to those reported for Hong Kong postgraduate nursing trainees \((\text{Kember et al., 2000})\), end-of-year 3 in practice Canadian nursing students \((\text{Lethbridge et al., 2011})\), and year 3 and year 5 UK medical students in a Problem Based Learning
(PBL) curriculum (Ambrose and Ker, 2014). A possible explanation for this could be the particular nature of KCLDI undergraduate programme - it is a research-led degree that stimulates critical thinking (as a process of reflection) and demands extensive understanding to be successful in an intensive learning environment which promotes early responsibility for patient care (Murray et al., 2003; Kings College London Dental Institute, 2013).

Clinical postgraduate and PhD trainees showed significantly higher average scores in Critical Reflection than undergraduate students (Figure 4-2). Similarly, the latter mean scores in Reflection were statistically different than those from PhD trainees. Thus, given that the levels of reflective habits from undergraduate students and postgraduate trainees are not the same, the third hypothesis of this Chapter cannot be accepted: The levels of reflective habits from undergraduate students and postgraduate trainees are the same.

These differences between students and postgraduate trainees reflective thoughts, might be explained by the professional practice model described by Schön (1983), in which unusual problems or cases, such as those possibly confronted by both clinical postgraduate and PhD trainees, might have evoked their critical reflection. Another plausible explanation might be the fact that both these groups exhibited a higher mean age at the time of the study (Table 4-2), which might indicate that they had several years of professional experience reflecting upon their professional practices (Boenink et al., 2004).

The small 0.2 mean score difference exhibited by PhD trainees between Reflection (18.3) and Critical Reflection (18.1) is noteworthy; this was 1.3 for clinical postgraduates and 3.0 for undergraduate students. PhD trainees are continuously challenged by evidence and new ideas they come across and possibly have more opportunities to develop their reflection and critical reflection skills. Similar minor differences between Reflection and Critical Reflection skills have been reported by Yuen Lie Lim (2011) in a study of 111 polytechnic students in the third year of a PBL Diploma (engineering, information technology, technology for the arts, and applied sciences), where means were 15.2 for Reflection and 15.3 for Critical Reflection.

Consistent with our results, previous studies using the same RQ (Phan, 2009) and other methodologies (Aronson et al., 2011; Carr and Johnson, 2013), found no gender difference in subjects’ reflective habits. However, Boenink et al. (2004), using
reflection-evoking case vignettes, found female medical students had slightly higher reflective scores than male medical students ($p < 0.018$). In contrast, in our study the $\geq 24$ years old and those with previous university degree groups showed statistically significant higher mean scores than their corresponding counterparts in their Reflection and Critical Reflection habits. This might be explained by the larger proportion of clinical postgraduate and PhD trainees in the senior and previous degree groups and the aging maturational effect (Sandars, 2009), with the concomitant development of critical thoughts and problem-solving skills distinctive of university education (Kember et al., 2000; Cano, 2005).

As explained by Yuen Lie Lim (2011), the positive correlations between Understanding, Reflection and Critical Reflection scales can be explained by the questionnaire’s theoretical bases that scales lie in a continuum (Lim, 2011), and contribute to an overall construct of reflective thinking (Kember et al., 2000). In this context, those students and trainees with high Reflection and Critical Reflection skills may also develop their Understanding abilities, particularly in more theoretical and less practical courses (Kember et al., 2000).

Reflection Questionnaire scores of Habitual Action, Understanding and Reflection from participating undergraduate students and their high stakes end-of-year examinations were positively correlated. Under these circumstances, the fourth hypothesis of this Chapter can be accepted: Undergraduate dental students’ reflective habits and their high stakes academic achievements are positively correlated.

Consequently, written essays on clinical scenarios, designed to examine application, appraisal, integration and interpretation of knowledge of BDS 4 and 5 students (Paper 1b in Table 4-7), were positively correlated to students’ Understanding and Reflection scores. This might indicate that essay cases with ill-defined problems required students not just to recall theoretical knowledge, but also to understand and apply it to practical situations in order to eventually solve the essay questions (Kember et al., 2008). Further, “reflection is the central dynamic involved in problem solving” (Mezirow, 1991), which explains why students with higher Reflection scores in our study performed better in this kind of examination.

In contrast, short answer and short notes questions designed to assess knowledge recall of BDS 1, 2 and 3 students (Paper 1a in Table 4-7) did not correlate to any of the RQ constructs. As confirmed by Boyd (2002), the initial years of a dental curriculum have a
large number of courses and students have little time to process and learn in a meaningful way. Further, similar results were reported in a large (N=581) and detailed study on second-year undergraduate university mathematics students from the Fiji Islands (Phan, 2008), where RQ scores did not relate to their assessment marks (not reflection encouraging tutorials, written assignments and final examination, at this stage).

Single best answer and extended matching multiple choice questions (Paper 2 in Table 4-7), aimed at testing the degree of understanding of a concept or topic from different aspects of the course by knowledge recall, association and classification of concepts, were positively correlated to Understanding. Likewise, 196 second year medical students’ end-of-term reflective skills correlated positively with their “cognitive” (no details of the assessment are given) academic achievements (Sobral, 2001). Further, the amount of reflective text written in a semi-structured portfolio of 91 fifth-year medical students positively correlated ($p<0.001$) with their success in a final examination comprising written case histories, multiple choice questions and essay (Lonka et al., 2001).

The case of the OSCE is quite particular as it revealed a positive correlation with Habitual Action as well as with Understanding. As a competency examination and in accordance with normal practice (Schoonheim-Klein et al., 2005), our students rotated between a series of 5 minute standardised stations where their understanding of a given clinical situation and their knowledge application to solve the presented clinical case was tested. To succeed at OSCE stations, students need to demonstrate that they have developed their understanding and decision making by learning to restrict themselves to the relevant features and aspects of the presented situation (Dreyfus et al., 1988; Dreyfus, 2004). Our results correspond with those in a study of year 4 medical students (Carr and Johnson, 2013), though a different questionnaire did not find a positive correlation between self-reflection and the OSCE. Further, the correlation between OSCE and Habitual Action might be explained by the observation of Schön (1983) that common cases – easy or expected OSCE stations, for example – could have possibly been approached by competent students in an habitual way as they were used to the presented scenarios, performing automatically with little conscious thinking. It should be borne in mind that despite Habitual Action being a non-reflective action, this response would be appropriate, for example, in emergency circumstances where actions are automatic without conscious thoughts, after which one can reflect about these
actions. However, it should not be relied on too often, especially in “complex and challenging situations, or to advance knowledge” (Lethbridge et al., 2011).

The Clinical Reasoning Examination (CRE), based on a similar assessment used in the Membership of the Joint Dental Faculties of The Royal College of Surgeons of England, was introduced at KCLDI for BDS year 5 Finals to replace the “single unseen case”. The aim is to assess communication and clinical reasoning skills in relation to clinical scenarios in the four main clinical areas of Child Dental Health, Acute Dental Care and Clinical Governance, Adult Dental Health and Clinical Diagnostic Sciences, and Oral Surgery. Thus, after 45 minutes studying four different clinical cases, students rotate between the same number of stations, each dedicated to one of the presented clinical cases, and are independently and structurally marked by two examiners. Students are marked on their level of competence in clinical reasoning of the case by articulating not only their deep knowledge and understanding, but also their ability to apply it to clinical problems (problem solving). These examination results and their positive correlation with Understanding and Reflection can again be supported by Schön’s observations (1983), as unusual problems or cases – e.g. clinical reasoning scenarios – and especially those with higher implications, lead to higher order thinking.

In line with this, Dreyfus et al. (1988) suggest deliberation should occur before action is taken when outcome performance is crucial and time permits. Further, in a recent study (Ambrose and Ker, 2014), reflection was identified as a recurrent learning mechanism in medical students and correlated with “knowledge of actions to take”, which, coincidently, is one of the assessment criteria of the CRE. Moreover, and according to Leung and Kember (2003), students with high Understanding and Reflection scores, as those good performers in the CRE and essays, employ a desirable deep approach to learning, which, as stated by Rolfe and McPherson (1995), is characterised by “the active pursuit of understanding and application of knowledge”, with the information being stored in the long-term memory (Harden and Laidlaw, 2013).

The Case Presentation examination is part of BDS Year 5 Finals, in which students present a clinical case they have treated in a poster format. Subsequently, two examiners check the actual treatment in the patient’s mouth and interact with the student on the diagnosis and rationale of the treatment provided, seeking to reveal the student’s level of understanding. It was expected that students’ scores from this examination would correlate with the Understanding scale of the RQ; however, this was not the case. A
possible explanation for this might be the effect of unrestricted access to staff during the planning and treatment phase of the clinical case, meaning that students were not required to think for themselves and hence not reflecting. Another likely reason for this could be that other end-of-year academic examinations’ demands occur simultaneously and might have relegated the Case Presentation to a less desirable surface approach (Cano, 2005).

Regardless of the positive correlations between Habitual Action, Understanding and Reflection scores with different examination methods, this was not seen with Critical Reflection. That is, students with high Critical Reflection scores did not perform better or worse in any end-of-year examinations. The reason for this might be found in the conceptual design of the RQ itself. Thus, the “Critical Reflection” dimension of Kember et al. (Kember et al., 2000), based on Mezirow’s “Premise Reflection” (Mezirow, 1991), are the highest levels of reflective thinking and imply the possibility for perspective transformation “guided by the identifying and judging of presuppositions” (Mezirow, 1991). This has been described as a difficult (Jonas-Dwyer et al., 2013), lengthy and painful process (Kember et al., 2000), associated with future intentions (Ambrose and Ker, 2014), and therefore not often observed (Lethbridge et al., 2011). A longitudinal assessment programme based on a meaningful experience interpretation (Mezirow, 1991), might help to ascertain a correlation with Critical Reflection.

The results of the present Chapter contribute to five aspects regarding dental students’ reflective habits that could prove useful for dental teaching and learning.

Firstly, and to the authors’ best knowledge, this is the first study which reports the assessment of dental undergraduate students’ and postgraduate trainees’ reflective habits following the four scales of the RQ protocol developed by Kember et al. (Kember et al., 2000). This is particularly important as this instrument has been utilised in a wide variety of disciplines which permits trustable comparisons of students’ reflective habits.

Secondly, knowing students’ scores in each of the RQ constructs could be helpful in investigating the effects of a given teaching and learning environment or a specific course or intervention on the reflective thinking of students.

Thirdly, some students are better reflectors than others, and thus, knowing individual student reflective patterns could be used to assist those less-able reflectors to develop
this skill and ideally improve their performance.

Fourthly, as reflective thinking did not show major distinctions between males and females, our attention in tutoring should be focused on younger students (≤ 23 years old) and those without a previous university degree, for they presented significantly lower Reflection and Critical Reflection scores.

Fifthly, a comprehensive assessment programme including a variety of assessment methods could stimulate more than a single construct of reflective thinking. Thus, whilst multiple choice questions encourage them to “understand”, ill-defined problems such as those from essays and clinical reasoning exercises would elevate this into the reflection scale.

However, the present Chapter study has some limitations. Probably the most evident is that it is only an observational study. As such, the results can only insinuate some possible roles for the Reflective Questionnaire. Thus, it would be desirable to have a study design allowing for pre- and post-observation reflection scores of similar groups during specific periods and examine the influence of different variables. Further, an intervention process could be included to examine its effect on students’ reflective skills and to investigate whether the RQ is sufficiently sensitive to detect changes as a result of the intervention. Another drawback is the absence of any kind of assessment method of clinical postgraduate and PhD trainees to correlate to their reflective habits.

4.7 Conclusions

The results demonstrate that dental students and postgraduate trainees’ non-reflective and reflective habits can be assessed reliably and validly using the self-reported Kember et al. Reflection Questionnaire (2000). Consequently, the long-standing tradition that the “collective wisdom” of the staff educators is an adequate source for evaluating students’ reasoning (Hendricson et al., 2006), can at least now be complemented with this reliable and valid questionnaire.

The studied Reflection Questionnaire, as a diagnostic instrument, not only allowed the determination of reflective and non-reflective skills of dental students, but also their correlation with different assessment methods. Thus, its use for a pre- and post-intervention evaluation of reflective skills, after prospectively using the peer-assessment and peer-feedback protocol, as it will be described in the next chapter, appears appropriate.
Chapter 5
Implementing a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback

Figure 5-1 Flowchart of the study’s research chapters. The fifth chapter is highlighted to help in providing an overall view.

5.1 Introduction
The clinical environment focuses students on actively learn from integrated real problems and affords teachers the opportunity to model professional thinking, behaviour and attitudes (Spencer, 2003). However, time pressures, staff competing demands and increasing numbers of students may adversely affect this (Spencer, 2003) (Chapter 1, Clinical workplace-based assessment and the role of peers in page 81). Current practice of WPBA has also been criticised for its absence of direct observation (Driessen and
Scheele, 2013), lack of informed and immediate feedback, difficulties in finding staff assessors (Quantrill and Tun, 2012) and limited opportunity for reflection and discussion (Spencer, 2003).

After the positive results of the peer-assessment protocol piloted in Chapter 3, with regards to the ability of undergraduate students to reliably peer-assess one another in all four GDC domains, and their perceived positive perceptions, it seems reasonable to think whether this protocol implemented in a longer period of time and in a larger sample of students, would help to overcome, at least partially, the current WPBA criticism described above. Probably most importantly, the protocol’s potential to foster students’ reflection and academic performance needs to be ascertained. This belief has been encouraged by the recent published study of Ali et al. (2014), who compared the ratings of a formative peer-assessment of second year undergraduate dental student’s clinical competence in simulated tooth extractions, with the ones of an equivalent, one week later, summative assessment scored by academic supervisors. Hence, in similar conditions to our protocol, students working in pairs alternated roles of assessor and trainee with their pre-clinical partner. After performing two tooth extractions in 60 minutes, the assessor acting student, following a given assessment criteria, graded and most importantly, provided verbal and written feedback to the trainee student. Despite results showing weak statistical correlations, 84% of students were awarded the same rating by peers and academic supervisors. Further, whilst a 5.7% who passed the peer-assessment later failed the summative one, a notable 10% improved their peer-assessment score in the later supervisor summative assessment. As the authors presented it, these latter results could be related to students learning from their formative peer-assessment experience.

Despite the Ali et al. (2014) study encouraging results for dental students’ peer-assessment, together with the ones presented in Chapter 3, it seems that peer-assessment in dental education as a method to contribute to each other’s learning process, is underexplored as no study has been published on the prospective use of peer-assessment and peer-feedback in dental undergraduate students, despite the potential dental peers have to contribute to each other’s learning process, as described in Chapter 1 (page 53).

Furthermore and from this “learning” point of view, a repetitive peer-observation, peer-assessment and provision of a constructive and immediate dialogic peer-feedback
(agreeing challenges to change performance) (Sargeant et al., 2011), might contribute to each other’s learning process encouraging self-reflection (Schön, 1983; Ashley et al., 2006; Boud and Falchikov, 2007; Miller et al., 2010) and critical skills (Dochy et al., 1999; Speyer et al., 2011), and by doing so improve their clinical performance (Boursicot et al., 2011).

5.2 Aim

Following a pilot study (Chapter 3), this Chapter aimed to evaluate the reliability, validity, feasibility, acceptability and educational impact of the same structured protocol of formative, prospective peer-assessment of undergraduate pre-clinical and clinical dental students’ skills as a framework for the provision of immediate peer-feedback. Further, the influence of the peer-assessment protocol on the students’ reflective skills was also examined.

5.3 Hypotheses

i. The implemented structured peer-assessment protocol is reliable.

ii. The implemented structured peer-assessment protocol is valid.

iii. The implemented structured peer-assessment protocol is feasible.

iv. The implemented structured peer-assessment protocol is acceptable.

v. The implemented structured peer-assessment protocol has a positive educational impact.

vi. The peer-assessment exercise stimulates students’ reflective skills.

5.4 Materials and Methods

5.4.1 Instruments for peer-assessment and feedback

The three peer-assessment instruments developed for this research (Chapter 3, page 110), were implemented in a larger sample and for a longer period of time.

Consequently, pre-clinical work-based peer-DOPS (Appendix 12), designed for peer-
assessment of any training procedure performed at the simulation skills laboratory, along with both clinical workplace-based peer-DOPS (Appendix 13), planned for the peer-assessment of whichever clinical procedure students performed on their patients, and the peer-mCEX (Appendix 14), meant for the peer-assessment of patients’ clinical examinations/assessment, were used as a framework for a continuous and structured peer-assessment and peer-feedback protocol of undergraduate pre-clinical and clinical dental students, respectively.

5.4.2 Instruments to assess academic achievement outcomes

In order to establish the possible effects of the studied peer-assessment and feedback protocol on students’ academic performance, their high-stakes end-of-year examination marks from the previous 2012 academic year (baseline) as well as those from the actual 2013 studied period (resultant), were collected and subsequently correlated to the peer-assessment variables.

Two student cohorts were studied:

a) BDS 2 baseline end-of-year examinations included written short answer questions (SAQs) and short note questions (SNQs) which constituted Paper 1, along with online multiple choice questions (MCQs), which formed Paper 2. The resultant end-of-year examinations comprised the same previous instrument formats (Paper 1 and Paper 2), supplemented by a hurdle clinical skills examination with an oral (viva) and practical components.

b) BDS 5 baseline end-of-year examinations involved essays on clinical scenarios (Paper 1), online MCQs (Paper 2), and an Objective Structured Clinical Examination (OSCE); resultant end-of-year examinations (Finals) encompassed all three former formats with the addition of a Clinical Reasoning Examination (CRE) and a Case Presentation.

5.4.3 Instrument for reflection skills assessment

The self-reported Reflection Questionnaire (Kember et al., 2000) (Chapter 4) was used to examine the influence of the peer-assessment and peer-feedback protocol on students’ reflective skills, both at the beginning (baseline) and end (resultant) of the 2012/2013 academic year.
5.4.4 Implementation and data collection

One researcher (JT) supervised the whole peer-assessment process. In June and September 2012 (baseline), all 309 BDS 2 (N=154) and BDS 5 (N=155) students enrolled at KCLDI received an electronic invitation from the Director of Education to voluntarily participate in the peer-assessment and peer-feedback exercise, which started by completing a Survey Monkey® online version of the self-reported Kember RQ.

The same researcher (JT) also delivered several 15-minute tutorials to give all pre-clinical BDS 2 Conservative Dentistry and all clinical BDS 5 Primary Dental Care students a detailed explanation of the study peer-assessment protocol, emphasizing the possible benefits of feedback and specifically exposing the findings of Hattie and Timperley (2007).

Those students who consented to participate in the peer-assessment exercise, received a 45-minute training and familiarisation session (delivered by JT) on observation, peer-assessment, peer-feedback, action plan and completion of the peer-DOPS and peer-mCEX instruments. They were also encouraged to undertake a brief self-reflection on the received feedback and action plan and record it in a personal reflection diary, immediately after each peer-assessment and feedback encounter.

In September 2012 and following a naturalistic design (Crossley et al., 2007; Boursicot et al., 2011) in that the selection of peer observers and procedures was natural and uncontrolled, BDS 2 students attending the Conservative Dentistry skills laboratory course organised themselves in pairs to work at neighbouring phantom heads. Similarly, BDS 5 clinical partners registered for Primary Dental Care, alternated the dentist/assistant roles randomly allocated each session. In both cohorts, each student acted as “observer” and “trainee” during the first half of the day and then switched roles during the second half of the day.

BDS 2 students customarily performed their own procedures while “observing” each peer’s pre-clinical work every 15 minutes (to avoid interfering with their own pre-clinical work). BDS 5 students performed their usual clinical activities in pairs so that the assistant student “observed” the dentist student while treating the patient together.

In each case, the observed procedure was then used to formatively assess each of the peer-DOPS and peer-mCEX domains, and act as a grounded framework to provide
informed written (in the actual peer-assessment form) and verbal terminal dialogic feedback (Walsh et al., 2009), as well as to agree an appropriate action plan to address any developmental needs, at the end each session. Finally, after signing the forms and placing them in a specially designed delivery box, students reflected on the feedback and action plan and noted their thoughts in a private reflection diary. In order to facilitate and stimulate the reviewing of their challenges and reflections, all feedback and action plans were uploaded (within 24 hours) to a password protected personal electronic peer-feedback and reflection diary as part of the King’s e-Learning and Teaching Service (KEATS) platform (Figure 5-2).

![KEATS (King's E-Learning and Teaching Service)](#)

**Figure 5-2 Example of a BDS 5 student’s peer-feedback diary designed to help them recall their peer-feedback and agreed challenges to stimulate self-reflection (the name of the student has been hidden).**
Participating students freely decided when to peer-assess each other; whether to submit the completed forms in the delivery box, the number of encounters to complete; whether to stop participating and ask for the return of their completed forms.

In order to investigate students’ and their tutors perceptions of the prospective peer-assessment exercise, during the last two weeks of peer-assessment they were asked to anonymously answer the following questions using a 5-point Likert-scale (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree):

To what extent do you agree that the peer-assessment and feedback protocol used in this study:

a) assessed areas that correspond to the activity in the pre-clinic/clinic?

b) was easy and straightforward to use?

c) could be introduced in the future to all students at King’s College London Dental Institute as part of their pre-clinical/clinical education?

d) was acceptable and fair?

e) has helped you to identify learning needs and to improve your performance?

Questions b) and e) were not presented to tutors. Both student and tutor questionnaires included a final free text box for any further comments / problems / suggestions.

Lastly, before the study ended (May 2013), a third and final invitation from the Director of Education was sent asking all BDS 2 and BDS 5 students to complete the end of the academic year (resultant) RQ.

5.4.5 Students’ stimulus

All students participating in the study were offered entry to a prize-draw for a Kindle Fire HD (one for each BDS year) and receive a “Research Participation Certificate” signed by the KCLDI Director of Education for their portfolio (which were delivered before the end of the academic year).

5.4.6 Data analysis

All collected data from completed peer-DOPS and peer-mCEX forms was used to evaluate the reliability, validity (content, criterion and construct), feasibility, acceptability and educational impact of the peer-assessment and feedback protocol. Furthermore, RQ scores were used to examine the influence of the peer-assessment protocol on the students’ reflective skills.
As in the piloting (Chapter 3), peer-assessment scores were assigned a numerical value from 1 to 6. Thus, the “Starting to develop” initial stage of ability was given a score 1, the “Show initial capability” a score 2 and so on until the highest “Show constant extremely good ability”, which was given a score 6.

Peer-DOPS and peer-mCEx forms data completed by students were manually digitised by one researcher (JT) into a spreadsheet, whilst RQ responses were electronically downloaded from Survey Monkey® into the same combined spreadsheet. Scores from both RQs (baseline and resultant) and peer-assessment were checked for normality assumptions using histogram and box plot before carrying out any parametric analysis.

Descriptive statistics using mean, standard deviation, percentage and range were used to summarise participants’ characteristics as well as the total number of observations, average of encounters per student, peer-assessed procedure categories, peer-DOPS and peer-mCEx scores and observation and feedback times. RQ traits were also analysed using descriptive statistics while its reliability was quantified by computing Cronbach’s Alpha coefficient.

A G-study (Generalizability study) (Cronbach and Shavelson, 2004) was used to determine the contribution that all relevant factors (trainee, occasion, items, and their interactions) made to the results, independently for each peer-assessment form, and thus reflect their reliability (Crossley et al., 2002a). Additionally, a D-study (Decision study) was used to estimate the effect of the number of peer-assessment encounters on the forms reliability. Thus, a G-study with a crossed three-facet (trainee × occasion × items) random-effects using items as a fix facet was used for pre-clinical peer-DOPS (assessor students were crossed within trainee students as they worked in fixed pairs), and a nested three-facet random-effects using items as a fix facet was used for clinical peer-DOPS and peer-mCEx (assessor students were most of the times nested (unique) within trainee students as they worked in random pairs).

Descriptive statistics were used to study the applicability and thus the content validity of both peer-DOPS and peer-mCEx items to be assessed by peers. Thus, the number of “unable to comment” or missing values for each item was investigated to determine whether they were assessing “valid” content. The presence of considerable “unable to comment” or missing values fields in a specific item would mean the item was not applicable. Further, peer-DOPS and peer-mCEx criterion validity was studied using a
Pearson correlation analysis between the mean peer-assessment score students were given for each of the different tools along the study period, with the mean of all their resultant high-stakes end-of-year examinations marks, which were assessing the same outcomes and therefore considered as the gold standard (Streiner and Norman, 2008).

Finally, construct validity of the peer-DOPS and peer-mCEX was assessed in four different ways by analysing:

a) the forms internal structure using Cronbach’s Alpha coefficients.

b) the homogeneity of all three scales through inter-item correlation and item-total correlations.

c) whether students were able to differentiate performance through one way analysis of Variance (ANOVA) among the 11 preclinical peer-DOPS items, 13 clinical peer-DOPS items, and 9 peer-mCEX items, separately.

d) whether students could detect improvement in peers’ performance over time through repeated measures ANOVA.

The feasibility of the peer-assessment protocol was studied by descriptive statistics analysing observation and feedback time reported by students in all completed peer-DOPS and peer-mCEX forms. This was complemented by students’ and tutors’ anonymous answers to both perception questions: “To what extent do you agree that the peer-assessment and feedback protocol used in this study was easy and straightforward to use?”, and “…could be introduced in the future to all students at King’s College London Dental Institute as part of their pre-clinical/clinical education?”

Further, the proportion of peer-DOPS and peer-mCEX forms delivered fully completed containing a feedback and action plan was also considered to study feasibility. Finally, following the UK Committee of Postgraduate Dental Deans & Directors (COPDEND) standards for Foundation Training dentists of one evaluation of performance per month (COPDEND, 2009), and as the present study lasted 8 months, the number of students who completed 8 or more peer-assessment forms was recorded.

Both students’ and tutors’ acceptability of the protocol was examined through descriptive statistics of the anonymous answers to the perception question “To what extent do you agree that the peer-assessment and feedback protocol used in this study was acceptable and fair?”
The educational impact of the peer-assessment protocol was studied through four different paths:

a) The peer-assessment and peer-feedback “effect size” on progress as described in Chapter 1 (Educational Impact in page 51) according to Hattie and Timperley (2007), was calculated for those students who exercised the peer-assessment protocol as a whole group and per the number of encounters they completed, as well as for those who did not take part in the protocol.

Thus, the difference between the resultant (2013) and the baseline (2012) (before anyone started to use the peer-DOPS and/or peer-mCEX tools), end-of-year examination marks, divided by a composite standard deviation, gave us the “effect size” ($d$) which helped to understand the impact of the intervention (peer-assessment and peer-feedback) over the academic year (Hattie, 2012).

b) As a supplement of the effect size, a paired-samples t-test was used to compare the baseline (2012) and resultant (2013) end-of-year examination marks of those students who followed the peer-assessment protocol throughout the academic year. The same comparison was also performed for those students who did not follow the protocol, as well as for the whole class.

c) An analysis using independent-samples t-test and effect size was performed comparing the students’ resultant (2013) mean end-of-year examination marks, and each of their high-stakes end-of-year Finals between those students who practiced the peer-assessment and peer-feedback protocol, with those who did otherwise (control group).

d) As a result of students’ anonymous answers to the perception question “To what extent do you agree that the peer-assessment and feedback protocol used in this study has helped you to identify learning needs and to improve your performance?” using descriptive statistics.

The influence of the peer-assessment and feedback protocol on the students’ reflective skills was studied, initially, by comparing baseline (2012) and resultant (2013) students’ RQ scores from those students who followed the peer-assessment protocol as well as from those who did not, using the effect size formula and a paired-samples t-test. Subsequently, and again calculating the effect size in addition to independent-samples t-test, the relationship between the number of peer-DOPS and/or peer-mCEX students completed with their resultant (2013) RQ scores, also was explored.
All analyses were carried out using SPSS® version 21, except for the generalizability coefficient which was calculated using the software EduG 6.1e.

5.5 Results

5.5.1 Participants

A total number of 157 volunteer students (102 females and 55 males, aged 20 to 42, mean=23.9, sd=3.3) participated in the study. The pre-clinical group comprised 71 BDS 2 students in total (46% of the class), 40 of whom agreed to participate in the peer-assessment exercise, while the remaining 31 BDS 2 students completed both baseline (2012) and resultant (2013) RQs but did not take part in the peer-assessment protocol. Further, the clinical group comprised 86 BDS 5 students (55% of the class), 68 of whom agreed to take part in the peer-assessment protocol, whereas the other 18 BDS 5 students completed both RQs but no peer-assessment forms.

Thus, those 108 students (40 BDS 2 and 68 BDS 5) who practiced the peer-assessment protocol became the study group, whilst those 49 (31 BDS 2 and 18 BDS 5) who only completed both RQs were considered the control group. At the end of the study, a total of 26 (65%) BDS 2 and 43 (63%) BDS 5 students from the study group, along with 9 tutors (4 pre-clinical and 5 clinical) completed the end-of-study anonymous perceptions’ questionnaire of the prospective peer-assessment protocol.

5.5.2 Peer-assessment general data

Students who participated in the peer-assessment exercise completed a total of 1169 peer-DOPS and peer-mCEX forms during the whole study period. More specifically, 40 participating BDS 2 students completed 516 pre-clinical peer-DOPS forms with a range of 1 to 25 and an average of 12.9 forms each student, whilst the 68 participating BDS 5 students filled in 576 clinical peer-DOPS and 77 peer-mCEX with a range of 1 to 27, and an average of 9.6 forms each. As a consequence of both groups wide range of peer-assessment encounters, participating students were divided in three groups according to the number of peer-DOPS and peer-mCEX they completed, as shown in Table 5-1.
Table 5-1 Number of BDS 2 and BDS 5 students according to the number of peer-assessment encounters completed.

<table>
<thead>
<tr>
<th>Group</th>
<th>BDS 2</th>
<th>BDS 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N. of students</td>
<td>N. of peer-DOPS</td>
</tr>
<tr>
<td>1-4 completed forms</td>
<td>10 (25%)</td>
<td>38 (7%)</td>
</tr>
<tr>
<td>5-9 completed forms</td>
<td>11 (27.5%)</td>
<td>96 (19%)</td>
</tr>
<tr>
<td>≥10 completed forms</td>
<td>19 (47.5%)</td>
<td>382 (74%)</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>516</td>
</tr>
</tbody>
</table>

All the differently grouped pre-clinical (BDS 2) and clinical (BDS 5) procedures students assessed each other, along with their own rated case complexity for each encounter, are shown in Table 5-2 and Table 5-3 respectively.

Table 5-2 Number of pre-clinical procedures BDS 2 students peer-assessed each other using peer-DOPS according to the perceived complexity assigned to each procedure.

<table>
<thead>
<tr>
<th>Pre-clinical procedures</th>
<th>Case complexity</th>
<th>BDS 2 peer-DOPS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Amalgam</td>
<td>2</td>
<td>47</td>
</tr>
<tr>
<td>Caries removal</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Cavity preparation</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Composite</td>
<td>17</td>
<td>145</td>
</tr>
<tr>
<td>Core restoration</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Fissure Sealant</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>Pulp cap</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Root canal treatment</td>
<td>11</td>
<td>102</td>
</tr>
<tr>
<td>Rubber dam placement</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Temporary restoration</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Direct veneer</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Wax-up tooth morphology</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>Total</td>
<td>66</td>
<td>344</td>
</tr>
</tbody>
</table>
Table 5-3 Number of clinical procedures BDS 5 students peer-assessed each other using peer-DOPS and peer-mCEX according to the perceived complexity assigned to each procedure.

<table>
<thead>
<tr>
<th>Clinical procedures</th>
<th>BDS 5 peer-DOPS and peer-mCEX</th>
<th>Case complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
<td>Moderate</td>
</tr>
<tr>
<td>Amalgam</td>
<td>8</td>
<td>16</td>
</tr>
<tr>
<td>Bridge</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>Composite</td>
<td>62</td>
<td>33</td>
</tr>
<tr>
<td>Crown</td>
<td>33</td>
<td>31</td>
</tr>
<tr>
<td>Denture</td>
<td>23</td>
<td>31</td>
</tr>
<tr>
<td>Examination</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Inlay</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Onlay</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Root canal treatment</td>
<td>3</td>
<td>59</td>
</tr>
<tr>
<td>Scaling and root planning</td>
<td>13</td>
<td>31</td>
</tr>
<tr>
<td>Teeth splinting</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Treatment planning</td>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>Veneer</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Unstated</td>
<td>8</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>170 (26.0%)</td>
<td>278 (42.6%)</td>
</tr>
</tbody>
</table>

The peer scores in both groups of students ranged from 1 to 6 and were normally distributed (mean=4.1, sd=1.5, median=4.0, mode=5). The mean and standard deviations of the various peer-DOPS and peer-mCEX assessment measures from the BDS 2 and BDS 5 groups are summarised in Table 5-4.

Table 5-4 Summary statistics (mean (sd)) of various measures of BDS 2 and BDS 5 peer-assessment forms.

<table>
<thead>
<tr>
<th>Variables</th>
<th>BDS 2 peer-DOPS</th>
<th>BDS 5 peer-DOPS</th>
<th>BDS 5 peer-mCEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of observations</td>
<td>516</td>
<td>576</td>
<td>77</td>
</tr>
<tr>
<td>Overall observation time (in min)</td>
<td>116 (79.8)</td>
<td>85 (38.1)</td>
<td>56 (32.9)</td>
</tr>
<tr>
<td>Overall feedback time (in min)</td>
<td>8 (3.7)</td>
<td>5 (2.9)</td>
<td>5 (1.9)</td>
</tr>
<tr>
<td>Overall score (1 to 6)</td>
<td>3.5 (1.3)</td>
<td>5.2 (0.6)</td>
<td>5.3 (0.5)</td>
</tr>
<tr>
<td>Utility of Giving Feedback (1 to 6)</td>
<td>3.7 (1.1)</td>
<td>4.1 (1.1)</td>
<td>4.4 (1.3)</td>
</tr>
<tr>
<td>Utility of Receiving Feedback (1 to 6)</td>
<td>3.7 (1.1)</td>
<td>5.4 (0.8)</td>
<td>5.4 (1.0)</td>
</tr>
</tbody>
</table>
5.5.3 Reflection Questionnaires general data

Baseline (2012) RQ scores from all 157 sampled students ranged from 1 to 5 and were normally distributed (mean=3.7, sd=1.0, median=4.0, mode=4). Cronbach’s Alpha internal reliability coefficient was 0.702. The adding of these baseline (2012) scores to conform each scale value, revealed a mean of 11.3 (sd=3.0) for Habitual Action, 17.5 (sd=1.7) for Understanding, 17.0 (sd=2.2) for Reflection, and 14.4 (sd=2.5) for Critical Reflection. Similarly, resultant (2013) RQ scores from the same 157 students showed the same range and comparable distribution (mean=3.8, sd=1.2, median=4.0, mode=4), whereas the Alpha coefficient was 0.731. The mean scale values this time were 11.4 (sd=3.6) for Habitual Action, 17.7 (sd=3.0) for Understanding, 17.5 (sd=2.9) for Reflection, and 14.8 (sd=3.1) for Critical Reflection.

5.5.4 Reliability of peer-DOPS and peer-mCEX

The absolute Generalizability coefficient for pre-clinical peer-DOPS forms was 0.724, while for the clinical peer-DOPS and peer-mCEX forms it was 0.787 and 0.795, respectively.

*Table 5-5 D study for different number of encounters for pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX.*

<table>
<thead>
<tr>
<th>Number of encounters</th>
<th>BDS 2 peer-DOPS G coefficient</th>
<th>BDS 5 peer-DOPS G coefficient</th>
<th>BDS 5 peer-mCEX G coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.563</td>
<td>0.601</td>
<td>0.632</td>
</tr>
<tr>
<td>2</td>
<td>0.574</td>
<td>0.624</td>
<td>0.641</td>
</tr>
<tr>
<td>3</td>
<td>0.592</td>
<td>0.631</td>
<td>0.655</td>
</tr>
<tr>
<td>4</td>
<td>0.641</td>
<td>0.653</td>
<td>0.672</td>
</tr>
<tr>
<td>5</td>
<td>0.667</td>
<td>0.661</td>
<td>0.689</td>
</tr>
<tr>
<td>6</td>
<td>0.679</td>
<td>0.672</td>
<td>0.699</td>
</tr>
<tr>
<td>7</td>
<td>0.690</td>
<td>0.687</td>
<td>0.721</td>
</tr>
<tr>
<td>8</td>
<td>0.705</td>
<td>0.711</td>
<td>0.735</td>
</tr>
<tr>
<td>9</td>
<td>0.719</td>
<td>0.746</td>
<td>0.747</td>
</tr>
<tr>
<td>10</td>
<td>0.722</td>
<td>0.756</td>
<td>0.769</td>
</tr>
<tr>
<td>11</td>
<td>0.728</td>
<td>0.756</td>
<td>0.775</td>
</tr>
<tr>
<td>12</td>
<td>0.725</td>
<td>0.766</td>
<td>0.778</td>
</tr>
<tr>
<td>13</td>
<td>0.727</td>
<td>0.769</td>
<td>0.789</td>
</tr>
<tr>
<td>14</td>
<td>0.726</td>
<td>0.772</td>
<td>0.792</td>
</tr>
<tr>
<td>15</td>
<td>0.724</td>
<td>0.787</td>
<td>0.795</td>
</tr>
</tbody>
</table>
The results of the D-study showing the G coefficient for each peer-assessment form and for the different number of encounters are shown in Table 5-5. The generally accepted G coefficient of ≥0.7 as sufficient for ‘low-stakes’ assessment situations (Beard et al., 2011) was reached with 8 pre-clinical and clinical peer-DOPS, and 6 peer-mCEX forms. Additionally, the variance component estimates for the pre-clinical peer-DOPS and clinical peer-DOPS and peer-mCEX tools are shown in Table 5-6.

Table 5-6 Variance components estimates for each peer-assessment form.

<table>
<thead>
<tr>
<th>Variance components</th>
<th>BDS 2 peer-DOPS</th>
<th>BDS 5 peer-DOPS</th>
<th>BDS 5 peer-mCEX</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of overall variance</td>
<td>Proportion of overall variance</td>
<td>Proportion of overall variance</td>
<td></td>
</tr>
<tr>
<td>Trainee (t)</td>
<td>61.8%</td>
<td>41.9%</td>
<td>45.3%</td>
</tr>
<tr>
<td>Occasion (o)</td>
<td>4%</td>
<td>0.5%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Item (i)</td>
<td>3%</td>
<td>2.0%</td>
<td>4.3%</td>
</tr>
<tr>
<td>to</td>
<td>11.4%</td>
<td>30.1%</td>
<td>20.5%</td>
</tr>
<tr>
<td>ti</td>
<td>1.4%</td>
<td>5.0%</td>
<td>2.7%</td>
</tr>
<tr>
<td>oi</td>
<td>1%</td>
<td>0%</td>
<td>0.1%</td>
</tr>
<tr>
<td>toi</td>
<td>17.4%</td>
<td>20.5%</td>
<td>25.3%</td>
</tr>
</tbody>
</table>

5.5.5 Validity of peer-DOPS and peer-mCEX

Content Validity

There were 49 (0.9%) scores marked as “unable to comment” or missing values among 5676 scores from the 516 completed pre-clinical peer-DOPS forms. From these, 19 (0.33%) corresponded to Item 4, and 20 (0.35%) to Item 11 (Table 5-7).

Similarly, there were 80 (1.1%) scores marked as “unable to comment” among 7488 scores from the 576 completed clinical peer-DOPS. The highest corresponded to “Administers effective analgesia or safe sedation” (Item 4) with 61 (0.81%) “unable to comment” selections (Table 5-8). This is probably due to procedures not needing analgesia, such as dentures (Table 5-3). Other “unable to comment” selections were Item 7 (0.09%), Item 8 (0.08%), and Item 2 (0.05%). A much lower score was evident in the 2 (0.29%) “unable to comment” selections among all 693 scores from the 77 completed peer-mCEX, both at Item 1 (Table 5-9). Altogether, “unable to comment” or missing values accounted for 0.74% of all scores.
Table 5-7 The mean and standard deviation (sd) for the 11 peer-DOPS items for pre-clinical peer-assessment of BDS 2 students (N=516 observations).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Peer-DOPS assessment items for BDS 2</th>
<th>Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Knowledge of technique and dental materials used for this procedure</td>
<td>3.2 (1.2)</td>
</tr>
<tr>
<td>2</td>
<td>Preparing for procedure according to taught protocol</td>
<td>3.6 (1.2)</td>
</tr>
<tr>
<td>3</td>
<td>Technical skills, manual dexterity and instruments handling</td>
<td>3.4 (1.2)</td>
</tr>
<tr>
<td>4</td>
<td>Working position and indirect vision</td>
<td>3.3 (1.2)</td>
</tr>
<tr>
<td>5</td>
<td>Following sequence and completing accurately all steps of the procedure</td>
<td>3.6 (1.3)</td>
</tr>
<tr>
<td>6</td>
<td>Observing aseptic technique/ Infection control and safe use of instruments</td>
<td>3.9 (1.3)</td>
</tr>
<tr>
<td>7</td>
<td>Seeking help where appropriate</td>
<td>3.7 (1.3)</td>
</tr>
<tr>
<td>8</td>
<td>Managing time/punctuality effectively</td>
<td>3.3 (1.4)</td>
</tr>
<tr>
<td>9</td>
<td>Supporting and communicating effectively with colleagues and tutors</td>
<td>3.8 (1.2)</td>
</tr>
<tr>
<td>10</td>
<td>Overall ability to perform procedure</td>
<td>3.1 (1.2)</td>
</tr>
<tr>
<td>11</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>3.2 (1.3)</td>
</tr>
</tbody>
</table>

Table 5-8 The mean and standard deviation (sd) for the 13 peer-DOPS items for clinical peer-assessment of BDS 5 students (N=576 observations).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Peer-DOPS assessment items for BDS 5</th>
<th>Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Demonstrates understanding of indications, dental materials, complications and technique of the procedure</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>2</td>
<td>Obtains informed consent after explaining procedure &amp; possible complications</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>3</td>
<td>Demonstrates appropriate preparation pre-procedure</td>
<td>5.3 (1.0)</td>
</tr>
<tr>
<td>4</td>
<td>Administers effective analgesia or safe sedation</td>
<td>5.6 (0.8)</td>
</tr>
<tr>
<td>5</td>
<td>Demonstrates appropriate technical ability in line with usual practice</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>6</td>
<td>Demonstrates aseptic technique/Infection control &amp; safe use of instruments &amp; sharps</td>
<td>5.4 (0.9)</td>
</tr>
<tr>
<td>7</td>
<td>Deals with unexpected events or seeks help when appropriate</td>
<td>5.4 (0.9)</td>
</tr>
<tr>
<td>8</td>
<td>Completes post procedure managements</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>9</td>
<td>Communication skills (patient &amp; team)</td>
<td>5.5 (0.8)</td>
</tr>
<tr>
<td>10</td>
<td>Organisation/efficiency and time management</td>
<td>5.3 (0.9)</td>
</tr>
<tr>
<td>11</td>
<td>Consideration of patient/professionalism</td>
<td>5.6 (0.7)</td>
</tr>
<tr>
<td>12</td>
<td>Overall ability to perform procedure</td>
<td>5.4 (0.8)</td>
</tr>
<tr>
<td>13</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>5.4 (0.8)</td>
</tr>
</tbody>
</table>
Table 5.9 The mean and standard deviation (sd) for the 9 peer-mCEX items for clinical peer-assessment of BDS 5 students (N=77 observations).

<table>
<thead>
<tr>
<th>Domain</th>
<th>Peer-mCEX assessment domains for BDS 5</th>
<th>Mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Interviewing/history taking skills</td>
<td>5.4 (0.6)</td>
</tr>
<tr>
<td>2</td>
<td>Physical examination skills</td>
<td>5.3 (0.7)</td>
</tr>
<tr>
<td>3</td>
<td>Diagnostic skills and underlying knowledge base</td>
<td>5.1 (0.8)</td>
</tr>
<tr>
<td>4</td>
<td>Communication and listening skills</td>
<td>5.4 (0.7)</td>
</tr>
<tr>
<td>5</td>
<td>Clinical judgment and decision making</td>
<td>5.3 (0.8)</td>
</tr>
<tr>
<td>6</td>
<td>Consideration for patient/professionalism</td>
<td>5.4 (0.7)</td>
</tr>
<tr>
<td>7</td>
<td>Organisation/efficiency and time management</td>
<td>5.2 (0.7)</td>
</tr>
<tr>
<td>8</td>
<td>Overall clinical competence</td>
<td>5.5 (0.5)</td>
</tr>
<tr>
<td>9</td>
<td>Does the trainee show insight into his/her performance?</td>
<td>5.6 (0.5)</td>
</tr>
</tbody>
</table>

Furthermore, 81% of BDS 2 and 94% of BDS 5 students anonymously stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study assessed areas that correspond to the activity in the pre-clinic/clinic”. Similarly, tutors’ perception of the protocol’s validity reached 88% (Table 5.10).

Table 5.10 Anonymous BDS 2 (N=26), BDS 5 (N=43) students’ and Tutors’ (N=9) perceptions (%) of the peer-assessment protocol, assessed at the end of the study.

<table>
<thead>
<tr>
<th>To what extent do you agree that the peer-assessment and feedback protocol used in this study:</th>
<th>Source</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Nor Agree nor disagree</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>…assessed areas that correspond to the activity in the pre-clinic/clinic?</td>
<td>BDS 2</td>
<td>0%</td>
<td>0%</td>
<td>19%</td>
<td>73%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0%</td>
<td>0%</td>
<td>6%</td>
<td>73%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>Tutor</td>
<td>0%</td>
<td>0%</td>
<td>11%</td>
<td>44%</td>
<td>44%</td>
</tr>
<tr>
<td>…was easy and straightforward to use?</td>
<td>BDS 2</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0%</td>
<td>3%</td>
<td>6%</td>
<td>64%</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>Tutor</td>
<td>0%</td>
<td>0%</td>
<td>11%</td>
<td>22%</td>
<td>78%</td>
</tr>
<tr>
<td>…could be introduced in the future to all students at the Dental Institute as part of their pre-clinical/clinical education?</td>
<td>BDS 2</td>
<td>0%</td>
<td>4%</td>
<td>31%</td>
<td>62%</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>3%</td>
<td>3%</td>
<td>24%</td>
<td>52%</td>
<td>18%</td>
</tr>
<tr>
<td></td>
<td>Tutor</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>22%</td>
<td>78%</td>
</tr>
<tr>
<td>…was acceptable and fair?</td>
<td>BDS 2</td>
<td>0%</td>
<td>4%</td>
<td>11%</td>
<td>62%</td>
<td>23%</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0%</td>
<td>3%</td>
<td>12%</td>
<td>64%</td>
<td>21%</td>
</tr>
<tr>
<td></td>
<td>Tutor</td>
<td>0%</td>
<td>0%</td>
<td>22%</td>
<td>11%</td>
<td>67%</td>
</tr>
<tr>
<td>…has helped you to identify learning needs and to improve your performance?</td>
<td>BDS 2</td>
<td>0%</td>
<td>8%</td>
<td>27%</td>
<td>58%</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td>BDS 5</td>
<td>0%</td>
<td>9%</td>
<td>18%</td>
<td>64%</td>
<td>9%</td>
</tr>
</tbody>
</table>
Criterion Validity

The mean peer-assessment score students were given for pre-clinical or clinical peer-DOPS and peer-mCEX during the study period were correlated individually to each student’s mean of all high-stakes end-of-year examination marks. Thus, BDS 2 and BDS 5 peer-DOPS scores from each participant showed positive and statistically significant correlation with their respective mean end-of-year examination mark ($r=0.583$, $p=0.0001$ for BDS 2 and $r=0.446$, $p=0.0001$ for BDS 5) (Table 5-11). However, despite peer-mCEX scores from the 17 participants students exhibited a positive correlation with their end-of-year examination mark, this was not statistically significant ($r=0.263$, $p=0.307$) (Table 5-11).

Table 5-11 Correlation between peer-assessment scores and the mean of all end-of-the-year examinations for each of the three tested instruments.

<table>
<thead>
<tr>
<th>Group of students</th>
<th>N</th>
<th>Mean peer-assessment score (sd)</th>
<th>Mean end-of-year examination mark (sd)</th>
<th>Correlation</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS 2 peer-DOPS</td>
<td>40</td>
<td>3.5 (1.3)</td>
<td>65.5 (5.0)</td>
<td>0.583</td>
<td>0.0001</td>
</tr>
<tr>
<td>BDS 5 peer-DOPS</td>
<td>68</td>
<td>5.2 (0.6)</td>
<td>62.2 (4.5)</td>
<td>0.446</td>
<td>0.0001</td>
</tr>
<tr>
<td>BDS 5 peer-mCEX</td>
<td>17</td>
<td>5.3 (0.5)</td>
<td>62.9 (2.1)</td>
<td>0.263</td>
<td>0.307</td>
</tr>
</tbody>
</table>

Construct Validity

The internal structure of all three peer-assessment instruments was assessed using Cronbach’s Alpha coefficients to indicate construct validity (Downing, 2003). Alpha was 0.964 for the 516 completed BDS 2 peer-DOPS forms, 0.960 for the 576 completed BDS 5 peer-DOPS forms, and 0.867 for the 77 completed BDS 5 peer-mCEX forms.

Further, pre-clinical peer-DOPS inter-item correlations range from a minimum of 0.571 to a maximum of 0.800. The highest item-total correlation (0.861) was for “Seeking help where appropriate” (Item 7), while Cronbach’s Alpha did not improve by deleting any item and only decreased by 0.005. Regarding clinical peer-DOPS, inter-item correlations range from a minimum of 0.422 to a maximum of 0.805 and the highest item-total correlation (0.852) was for “Overall ability to perform procedure” (Item 12). Likewise BDS 2, clinical peer-DOPS Cronbach’s Alpha did not improve by deleting any item. Somewhat differently, peer-mCEX inter-item correlations range from a lower minimum of 0.082 to a maximum of 0.784 and the highest item-total correlation (0.798)
was for “Overall clinical competence” (Item 8). Cronbach’s Alpha only improved to 0.881 by deleting Item 3 “Diagnostic skills and underlying knowledge”.

A third way of analysing peer-DOPS and peer-mCEX construct validity was by testing the students’ ability to detect differences in performance among the different peer-assessment domains. Thus, the scores from each of the eleven BDS 2 peer-DOPS assessed domains (Table 5-7) showed a statistically significant variance (F=18.232, \( p<0.0001 \)). Although at a lower level, BDS 5 peer-DOPS thirteen domains (Table 5-8) also differed significantly (F=2.593, \( p=0.002 \)), as did the peer-mCEX (Table 5-9) nine domains (F=2.027, \( p=0.045 \)).

![Figure 5-3 Overall mean peer-DOPS and peer-mCEX scores and standard deviations by session for BDS 2 and BDS 5 cohorts.](image)

Students’ capability to perceive differences in performance over time using peer-DOPS and peer-mCEX was also examined. As a result, the fortnight trends in performance scores showed a progression over the academic year (Figure 5-3) which was statistically significant for BDS 2 peer-DOPS (F=1528.652, \( p<0.0001 \)), as well as for BDS 5 peer-DOPS (F=135.335, \( p<0.0001 \)) and peer-mCEX (F=837.141, \( p<0.0001 \)). However, the
profile was uneven, especially between BDS 2 and BDS 5 scores. Thus, the former never reached the range of scores between 5 and 6 that was attained by BDS 5 peer-DOPS and peer-mCEX as early as the end of November. However, BDS 2 scores showed an almost permanent upward trend from January, which continued rising until the last encounter. Conversely, BDS 5 scores showed a steady state from November onwards.

5.5.6 Feasibility of peer-DOPS and peer-mCEX

The mean observation time (Table 5-4) for all BDS 2 peer-DOPS encounters was 116 minutes with some variability (sd=79.8). This indicates that they “observed” each other for almost the whole pre-clinical session of 180 minutes. In contrast, BDS 5 clinical students are required to see at least two patients every half day. Thus, the mean of 85 minutes in which they observed each other while performing a clinical procedure on their patients (recorded at peer-DOPS) is consistent with the 90 minute assigned sessions. The time it took for the same BDS 5 clinical students to clinically perform an examinations/assessment of a patient registered on the peer-mCEX forms (56 minutes) was less than that when treating patients. Regarding feedback time (Table 5-4), 46.7% of BDS 2 encounters were performed within the suggested 5 minutes (Norcini and Burch, 2007), while this rose to 74.9% for BDS 5 encounters. However, better evidence for this is the 100% of BDS 2 and 91% of BDS 5 students who stated they agree or strongly agree “that the peer-assessment and feedback protocol used in this study was easy and straightforward to use”. Furthermore, 66% of BDS 2 and 70% of BDS 5 students and also 100% of tutors, stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study could be introduced in the future to all students at King’s College London Dental Institute as part of their pre-clinical/clinical education” (Table 5-10).

The total of peer-DOPS and peer-mCEX delivered fully completed with feedback and action plan reached a total of 1104/1169 (94.4%) forms. Despite being fully completed in all other items, a total of 40 (3.4%) forms were submitted in the delivery box without written feedback comments, while 25 (2.1%) did not contain transcribed agreed action plans. Finally, the number of students who completed 8 or more peer-assessment forms (at least one per month)(COPDEND, 2006) in BDS 2 was 27/40 (67.5%) whereas in BDS 5 this number was 29/68 (42.6%). Overall, these results indicate the studied protocol is feasible to be implemented in dental undergraduate pre-clinical and clinical
courses and therefore the third hypothesis of the Chapter can be accepted: The implemented structured peer-assessment protocol is feasible.

5.5.7 Acceptability of peer-DOPS and peer-mCEX

Students who anonymously stated that they agree or strongly agree “that the peer-assessment and feedback protocol used in this study was acceptable and fair” reached 85% in both BDS 2 and BDS 5. Likewise, tutors’ perception of the protocol’s acceptability and fairness amounted to 78% (Table 5-10). Selected BDS 2 and BDS 5 students’ free text feedback comments extracted from their perceptions questionnaires are shown in Table 5-12.

Table 5-12 Selected anonymous students’ and tutors’ free text comments on the peer-assessment protocol.

<table>
<thead>
<tr>
<th>Student / Tutor</th>
<th>Number of encounters</th>
<th>Feedback</th>
<th>Acceptability score</th>
</tr>
</thead>
<tbody>
<tr>
<td>BDS 2</td>
<td>&gt;13</td>
<td>From my experience of this process I believe giving the students a structured method to review/analyse their work has been extremely useful.</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>BDS 2</td>
<td>10-13</td>
<td>Think that these forms should be filled out by tutors plus observing student.</td>
<td>Agree</td>
</tr>
<tr>
<td>BDS 2</td>
<td>10-13</td>
<td>It is dull and time consuming. Clinical partner can just tell me.</td>
<td>Nor Agree nor disagree</td>
</tr>
<tr>
<td>BDS 2</td>
<td>&gt;13</td>
<td>I always felt rushed. Main problem was having enough time to fill in the forms at the end of each session. There should be time set aside for it.</td>
<td>Agree</td>
</tr>
<tr>
<td>BDS 2</td>
<td>6-9</td>
<td>The feedback was fairly useful.</td>
<td>Agree</td>
</tr>
<tr>
<td>BDS 2</td>
<td>10-13</td>
<td>More concise form. Useful but does take too long to complete.</td>
<td>Nor Agree nor disagree</td>
</tr>
<tr>
<td>BDS 5</td>
<td>10-13</td>
<td>Reflection is very important in clinical improvement. iPad/tablet form would be easier to track and feedback on.</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>BDS 5</td>
<td>6-9</td>
<td>These forms should be introduced on all the clinics. Very useful and helpful for us. Students are always asking for feedback on performance.</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>BDS 5</td>
<td>5-3</td>
<td>I feel that this exercise would be more useful for BDS 1-3 students</td>
<td>Agree</td>
</tr>
<tr>
<td>BDS 5</td>
<td>2-5</td>
<td>I already reflect on what went well / what could be improved after each session but don’t write it down.</td>
<td>Agree</td>
</tr>
<tr>
<td>Tutor</td>
<td>-</td>
<td>The main reason for the success of this very important project lies with the consistent attendance of the investigator who directed his research in a non-intrusive manner and ignited the interests of all who took part.</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>Tutor</td>
<td>-</td>
<td>Yes, they all need feedback on results</td>
<td>Agree</td>
</tr>
<tr>
<td>Tutor</td>
<td>-</td>
<td>Peer review and reflective conversation resulted in self-criticism and improvement of personal and clinical skills. It points out problems that they may not be aware of. Strongly recommend it.</td>
<td>Strongly agree</td>
</tr>
</tbody>
</table>
5.5.8 Educational impact

The first and second paths to assess the peer-assessment and peer-feedback protocol’s educational impact analysed the effect size and the statistical significance of the differences between students’ baseline (2012) end-of-year examination marks (before the use peer-DOPS and/or peer-mCEx tools), and the resultant (2013) end-of-year examination marks. This was performed for BDS 2 students (Table 5-13) comparing the mean of all high-stakes examinations as well as for those specific tests with the same format at both times, namely written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions). An initial comparison between baseline scores of all BDS 2 studied groups using one way ANOVA test, showed the observable small differences between mean scores were not statistically significant for the average of all examination (F=0.232, p=0.948), written SAQs and SNQs (F=1.584, p=0.165), and online MCQs (F=1.928, p=0.090). This indicates all studied groups started the intervention in similar conditions.

Subsequently, the paired t-test comparing BDS 2 baseline and resultant mean scores for all groups showed no statistically significant differences, despite an observable tendency for those 19 students who completed 10 or more peer-assessment and peer-feedback encounters to show higher resultant scores (≥3.3 points) (Table 5-13). In line with this, the effect size (d) on these 19 students was higher than the “gold standard” d=0.40 average schooling effect (Hattie and Timperley, 2007; Norcini and Burch, 2007), in all three cases: average of all examination (d=0.59), written SAQs and SNQs (d=0.67), and online MCQs (d=0.42). Further, those 114 students from the same class who did not take part in the peer-assessment and peer-feedback protocol, showed no statistically significant differences and small or even negative effect sizes (Table 5-13).

The same analyses was performed for BDS 5 students comparing the effect size and statistical significance of the differences between baseline and resultant means of all high-stakes end-of-year examination marks along with those specific tests using the same format at baseline and resultant observation times (Table 5-14). These were, essay on clinical scenarios, online SAQs, SNQs and MCQs, and OSCE. As with BDS 2 results, the comparison between baseline scores from all BDS 5 groups using one way ANOVA test, did not reveal statistically significant differences for the average of all examination (F=0.783, p=0.562), essay on clinical scenarios (F=1.051, p=0.388), online SAQs, SNQs and MCQs (F=0.838, p=0.524), and OSCE (F=1.648, p=0.147).
However and in contrast to BDS 2, the paired t-test comparing BDS 5 baseline and resultant mean scores showed some statistically significant differences for some groups (Table 5-14). Thus, the difference between baseline and resultant average of all examinations from those 22 students who completed 10 or more peer-assessment and peer-feedback encounters was not only statistically significant ($p=0.009$) but with high positive effect size ($d=0.74$). Even more, the effect size for those 18 students who completed between 5 and 9 encounters was also higher than the 0.40 average ($d=0.46$), though the paired t-test showed no significance.

As depicted in the same Table 5-14, it seems that the peer-assessment protocol did not have any especial effect on students’ essay and OSCE marks, as all groups significantly increased and decreased their scores, respectively. However, this might also be due to other issues like for example, examination difficulty. The online SAQs, SNQa and MCQs only showed a small size effect ($d=0.33$) between baseline and resultant observations, though not statistically significant ($p=0.279$).

The third path to study the protocol’s educational impact studied the effect size and statistical significance (independent-samples t-test) of the differences between resultant (2013) end-of-year examinations from those students who practice the peer-assessment protocol against those who did not. Accordingly, BDS 2 students (Table 5-15) showed that the difference in the average of all examinations between those 40 students who practiced the peer-assessment protocol (mean=65.5, sd=5.0) and those 114 who did not (mean=63.1, sd=4.9) was statistically significant ($p=0.008$) as well as with a positive effect size ($d=0.48$). More specifically, this was led by a powerful effect size of $d=0.80$ and highly significant ($p=0.001$) difference between the group of those 19 students who peer-assessed each other 10 or more times with a mean of 67.1 (sd=3.8) in their average of all examinations, and those 114 students who did not peer-assess one another (mean=63.1, sd=4.9). Furthermore, the differences between these two group scores were statistically significant and with a high positive effect size in their online MCQs ($p=0.047$, $d=0.49$), Clinical Skills Examination viva ($p=0.050$, $d=0.48$) and practical ($p=0.021$, $d=0.56$). However, there was no measurable effect on written SAQs and SNQs for any of the studied groups.

A similar outline was shown by BDS 5 students’ results when comparing their Finals scores (Table 5-16). Thus, those 68 students who practiced the peer-assessment protocol showed a statistically significant difference ($p=0.028$) and a positive effect size ($d=0.36$)
in their average of all examination (mean=62.2, sd=4.5) when compared to those 87 students of the same class who did not take part in the peer-assessment study (mean=60.7, sd=3.7).

Once again, those 22 students’ scores who peer-assessed each other 10 or more times exhibited highly statistically significant differences and powerful positive effect sizes when compared to the 87 students who did not participate in the peer-assessment protocol, in their average of all examination ($p=0.0001, d=0.74$), essays on clinical scenarios ($p=0.0001, d=0.89$), online MCQs ($p=0.0004, d=0.80$), and Clinical Reasoning Examination ($p=0.011, d=0.59$). However, this impact was not observed comparing their OSCEs ($p=0.299, d=0.24$) and Case Presentation’s ($p=0.272, d=0.25$) scores as they obtained similar results.

Finally, a majority of 66% of BDS 2 and 73% BDS 5 students agreed or strongly agreed that the peer-assessment and peer-feedback protocol used in this study helped them to identify learning needs and to improve their performance (Table 5-10).
Table 5-13 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students’ average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were written SAQs (short answer questions) and SNQs (short note questions), and online MCQs (multiple choice questions).

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Average all examinations</th>
<th>Written SAQs, SNQs</th>
<th>Online MCQs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd)</td>
<td>p</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>baseline vs. resultant</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All class (baseline)</td>
<td>154</td>
<td>63.5 (9.7)</td>
<td>0.799</td>
<td>0.03</td>
</tr>
<tr>
<td>All class (resultant)</td>
<td></td>
<td>63.7 (5.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participating students (baseline)</td>
<td>40</td>
<td>62.5 (9.3)</td>
<td>0.087</td>
<td>0.41</td>
</tr>
<tr>
<td>All participating students (resultant)</td>
<td></td>
<td>65.5 (5.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-4 completed forms (baseline)</td>
<td>10</td>
<td>64.6 (8.8)</td>
<td>0.805</td>
<td>-0.13</td>
</tr>
<tr>
<td>1-4 completed forms (resultant)</td>
<td></td>
<td>63.8 (5.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-9 completed forms (baseline)</td>
<td>11</td>
<td>61.3 (10.8)</td>
<td>0.430</td>
<td>0.37</td>
</tr>
<tr>
<td>5-9 completed forms (resultant)</td>
<td></td>
<td>64.4 (6.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥10 completed forms (baseline)</td>
<td>19</td>
<td>63.4 (8.3)</td>
<td>0.094</td>
<td>0.59</td>
</tr>
<tr>
<td>≥10 completed forms (resultant)</td>
<td></td>
<td>67.1 (3.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No peer-assessment</td>
<td>114</td>
<td>63.9 (9.8)</td>
<td>0.478</td>
<td>-0.10</td>
</tr>
<tr>
<td>All non-participating students (baseline)</td>
<td></td>
<td>63.1 (4.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All non-participating students (resultant)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5-14 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and effect size (d), for the students’ average of all end-of-year examination marks as well as specifically for those examinations with the same format both at baseline and resultant observations, according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed. The latter examinations were essays on clinical scenarios, online SAQs (short answer questions), SNQs (short note questions) and online MCQs (multiple choice questions), and the OSCE (Objective Structured Clinical Examination).

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Average all examinations</th>
<th>Essay on clinical scenarios</th>
<th>Online SAQs, SNQs, MCQs</th>
<th>OSCE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>mean (sd) p d baseline vs. resultant</td>
<td>mean (sd) p d baseline vs. resultant</td>
<td>mean (sd) p d baseline vs. resultant</td>
<td>mean (sd) p d baseline vs. resultant</td>
</tr>
<tr>
<td>All class (baseline)</td>
<td>155</td>
<td>61.3 (4.7) 0.878 0.02</td>
<td>57.7 (5.0) 0.0001 0.69</td>
<td>63.1 (6.3) 0.744 -0.04</td>
<td>61.2 (6.3) 0.0001 -0.89</td>
</tr>
<tr>
<td>All class (resultant)</td>
<td>61.3 (4.2)</td>
<td>61.3 (5.0)</td>
<td>61.3 (5.0)</td>
<td>61.3 (5.0)</td>
<td>56.6 (4.0)</td>
</tr>
<tr>
<td>Peer-assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participating students (baseline)</td>
<td>68</td>
<td>60.8 (4.1) 0.059 0.33</td>
<td>57.1 (4.5) 0.0001 0.94</td>
<td>64.0 (6.2) 0.928 -0.02</td>
<td>62.4 (6.6) 0.0001 -0.98</td>
</tr>
<tr>
<td>All participating students (resultant)</td>
<td>62.2 (4.5)</td>
<td>62.2 (5.2)</td>
<td>62.2 (5.2)</td>
<td>62.2 (5.2)</td>
<td>56.9 (4.3)</td>
</tr>
<tr>
<td>1-4 completed forms (baseline)</td>
<td>28</td>
<td>60.0 (4.1) 0.945 -0.02</td>
<td>56.1 (3.0) 0.002 0.80</td>
<td>63.5 (5.8) 0.312 -0.28</td>
<td>60.8 (6.5) 0.0006 -1.03</td>
</tr>
<tr>
<td>1-4 completed forms (resultant)</td>
<td>59.9 (3.9)</td>
<td>59.9 (5.4)</td>
<td>59.9 (5.4)</td>
<td>59.9 (5.4)</td>
<td>55.6 (3.7)</td>
</tr>
<tr>
<td>5-9 completed forms (baseline)</td>
<td>18</td>
<td>60.5 (4.6) 0.232 0.46</td>
<td>57.3 (2.9) 0.0007 1.19</td>
<td>64.3 (6.6) 0.921 -0.04</td>
<td>62.1 (6.6) 0.009 -1.08</td>
</tr>
<tr>
<td>5-9 completed forms (resultant)</td>
<td>62.4 (3.3)</td>
<td>61.6 (2.9)</td>
<td>61.6 (2.9)</td>
<td>61.6 (2.9)</td>
<td>56.2 (4.3)</td>
</tr>
<tr>
<td>≥10 completed forms (baseline)</td>
<td>22</td>
<td>61.6 (4.1) 0.009 0.74</td>
<td>58.0 (6.2) 0.0001 1.10</td>
<td>64.5 (6.5) 0.279 0.33</td>
<td>64.0 (6.5) 0.0005 -1.12</td>
</tr>
<tr>
<td>≥10 completed forms (resultant)</td>
<td>64.9 (4.0)</td>
<td>65.2 (4.7)</td>
<td>65.2 (4.7)</td>
<td>65.2 (4.7)</td>
<td>58.1 (4.1)</td>
</tr>
<tr>
<td>No peer-assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All non-participating students (baseline)</td>
<td>87</td>
<td>60.5 (4.7) 0.728 0.05</td>
<td>58.3 (5.3) 0.002 0.47</td>
<td>62.3 (6.4) 0.753 -0.05</td>
<td>60.3 (6.0) 0.0001 -0.63</td>
</tr>
<tr>
<td>All non-participating students (resultant)</td>
<td>60.7 (3.7)</td>
<td>60.7 (4.7)</td>
<td>60.7 (4.7)</td>
<td>60.7 (4.7)</td>
<td>57.0 (4.5)</td>
</tr>
</tbody>
</table>
Table 5-15 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 114 non-participating students (No peer-assessment), for each of the students’ end-of-year examination marks (SAQs (short answer questions), SNQs (short note questions), MCQ (multiple choice questions) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

<table>
<thead>
<tr>
<th>BDS 2 end-of-year examination marks 2013</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Studied groups</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>All class</td>
</tr>
<tr>
<td>Peer-assessment</td>
</tr>
<tr>
<td>All participating students</td>
</tr>
<tr>
<td>1-4 completed forms</td>
</tr>
<tr>
<td>5-9 completed forms</td>
</tr>
<tr>
<td>≥10 completed forms</td>
</tr>
<tr>
<td>No peer-assessment</td>
</tr>
<tr>
<td>All non-participating students</td>
</tr>
</tbody>
</table>
Table 5.16 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (independent-samples t-test), and effect size (d) when comparing all groups to the 87 non-participating students (No peer-assessment), for each of the students’ end-of-year examination marks (SAQs=short answer questions, SNQs=short note questions, MCQs=multiple choice questions, OSCE=Objective Structured Clinical Examination) according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Average all examinations</th>
<th>Essay on clinical scenarios</th>
<th>Online MCQs</th>
<th>OSCE</th>
<th>Clinical Reasoning Examination</th>
<th>Case Presentation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd) p d diff. with no peer-assessment group</td>
<td>Mean (sd) d diff. with no peer-assessment group</td>
<td>Mean (sd) p d diff. with no peer-assessment group</td>
<td>Mean (sd) p d diff. with no peer-assessment group</td>
<td>Mean (sd) p d diff. with no peer-assessment group</td>
<td>Mean (sd) p d diff. with no peer-assessment group</td>
</tr>
<tr>
<td>All class</td>
<td>155</td>
<td>61.3 (4.2) 0.279 0.15</td>
<td>61.3 (5.0) 0.326 0.13</td>
<td>62.8 (5.3) 0.250 0.16</td>
<td>56.6 (4.0) 0.463 -0.10</td>
<td>64.3 (8.5) 0.432 0.11</td>
<td>61.7 (8.4) 0.424 0.11</td>
</tr>
<tr>
<td>All participating students</td>
<td>68</td>
<td>62.2 (4.5) 0.028 0.36</td>
<td>62.2 (5.2) 0.066 0.30</td>
<td>63.9 (4.4) 0.036 0.36</td>
<td>56.9 (4.3) 0.873 -0.03</td>
<td>65.6 (8.9) 0.152 0.24</td>
<td>63.0 (8.5) 0.106 0.27</td>
</tr>
<tr>
<td>1-4 completed forms</td>
<td>28</td>
<td>59.9 (3.9) 0.319 -0.22</td>
<td>59.9 (5.4) 0.436 -0.17</td>
<td>61.9 (6.1) 0.887 -0.03</td>
<td>55.6 (3.7) 0.125 -0.34</td>
<td>62.4 (6.9) 0.436 -0.17</td>
<td>60.6 (8.0) 0.890 -0.03</td>
</tr>
<tr>
<td>5-9 completed forms</td>
<td>18</td>
<td>62.4 (3.3) 0.128 0.45</td>
<td>61.6 (2.9) 0.502 0.19</td>
<td>64.1 (4.6) 0.172 0.40</td>
<td>56.2 (4.3) 0.530 -0.18</td>
<td>66.4 (7.0) 0.207 0.37</td>
<td>65.0 (8.3) 0.079 0.50</td>
</tr>
<tr>
<td>≥10 completed forms</td>
<td>22</td>
<td>64.9 (4.0) 0.0001 0.74</td>
<td>65.2 (4.7) 0.0001 0.89</td>
<td>66.2 (3.9) 0.0004 0.80</td>
<td>58.1 (4.1) 0.299 0.24</td>
<td>68.3 (8.3) 0.011 0.59</td>
<td>62.9 (8.6) 0.272 0.25</td>
</tr>
<tr>
<td>No peer-assessment</td>
<td>87</td>
<td>60.7 (3.7) - -</td>
<td>60.7 (4.7) - -</td>
<td>62.1 (5.1) - -</td>
<td>57.0 (4.5) - -</td>
<td>63.7 (7.6) - -</td>
<td>60.8 (8.2) - -</td>
</tr>
</tbody>
</table>

BDS 5 end-of-year examination marks 2013
5.5.9 Peer-assessment influence on students’ reflective skills

The influence of the peer-assessment protocol was also assessed by studying the relationship between baseline (2012) and resultant (2013) students’ reflection skills scores from those students who followed the peer-assessment protocol and those who did not. As described in Chapter 4, once completing the RQ, each participant received a score ranging from a minimum of 4 to a maximum of 20 for each Habitual Action, Understanding, Reflection and Critical Reflection scale. Consequently, the higher the score, the more agreement with engaging in the particular dimension that each scale measures. All groups showed higher scores for Understanding and Reflection scales both before (baseline) and after (resultant) the peer-assessment exercise, indicating that students employed these levels of thinking more than Habitual Action and Critical Reflection (Kember et al., 2000; Lethbridge et al., 2011; Chelliah and Arumugam, 2012). It is worth noting that the RQ assessment was identical on both occasions and hence the baseline and resultant scores therefore evaluated the real progress or regression in each of its four scales.

For BDS 2, statistically significant higher scores and positive effect sizes between baseline and resultant observations (paired t-test) were perceived in those 40 students who performed the peer-assessment exercise in Understanding \((p=0.003, d=0.65)\) and Reflection \((p=0.004, d=0.63)\) scales (Table 5-17). Once more, those 19 students who completed 10 or more peer-assessment encounters showed statistically significant differences and positive effect sizes in their baseline and resultant Understanding \((p=0.034, d=0.82)\) and Reflection \((p=0.008, d=0.95)\) scores. Additionally, and though not statistically significant, those students who completed between 5 and 9 forms also demonstrated a high effect size in Understanding \((d=0.77)\) and Reflection \((d=0.52)\).

In contrast, BDS 5 student’s results (Table 5-18) showed a statistically significant increase in their Critical Reflection skills from baseline to resultant \((p=0.003)\) only for those 22 students who completed 10 or more peer-DOPS and/or peer-mCEX encounters. The same group also demonstrated a positive effect size in their Critical Reflection skills \((d=1.04)\) as did those 18 students with 5 and 9 encounters \((d=0.46)\). A strong negative effect size was evident in the Habitual Action skills of those students with 10 or more peer-assessment encounters \((d=-0.65)\).

The BDS 2 independent-samples t-test and effect size of the differences between
resultant RQ scores from those students who exercised the peer-assessment and peer-feedback protocol, and those who completed the RQ but did not take part in the peer-assessment protocol, showed statistically significant results and positive effect sizes in Understanding and Reflection scales (Table 5-19). Thus, those 11 students with 5 to 9 peer encounters showed statistically significant higher \( (p=0.047) \) Understanding scores and a positive effect size \( (d=0.68) \), when compared to those 31 students who did not peer-assessed each other. Similarly, those 19 students with 10 or more encounters also showed significantly higher \( (p=0.011) \) Understanding scores and positive effect size \( (d=0.71) \) than those 31 who did none. Interestingly, the difference in Habitual Action between the 40 students who did participate in the peer-assessment protocol and the 31 who did not, revealed a high negative effect size \( (d= -0.46, \ p=0.05) \).

Finally, BDS 5 students who completed 10 or more peer-assessment encounters (\( N=22 \)) demonstrated a statistically significant higher Critical Reflection score \( (p=0.005) \) as compared to the group of 18 students who did not peer-assessed one another (Table 5-20). This difference also showed a large effect size of \( d=1.41 \). However, the same groups displayed a non-statistically significant difference in their Reflection score \( (p=0.826) \), despite had a high effect size \( (d=0.78) \).
Table 5-17 BDS 2 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants’ academic year and whether they participated in the peer-assessment exercise.

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd)</td>
<td>p</td>
<td>d</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>baseline vs. resultant</td>
<td></td>
<td></td>
<td>baseline vs. resultant</td>
</tr>
<tr>
<td>All who completed the RQ (baseline)</td>
<td>71</td>
<td>9.9 (2.3)</td>
<td>0.103</td>
<td>0.27</td>
<td>17.8 (1.2)</td>
</tr>
<tr>
<td>All who completed the RQ (resultant)</td>
<td>10.6 (2.8)</td>
<td>0.759</td>
<td>-0.05</td>
<td>17.8 (1.3)</td>
<td>0.003</td>
</tr>
<tr>
<td>Peer-assessment</td>
<td></td>
<td>9.9 (2.4)</td>
<td>0.643</td>
<td>-0.22</td>
<td>17.8 (1.0)</td>
</tr>
<tr>
<td>With RQ (baseline)</td>
<td>40</td>
<td>10.0 (2.7)</td>
<td>0.794</td>
<td>0.12</td>
<td>17.9 (1.4)</td>
</tr>
<tr>
<td>With RQ (resultant)</td>
<td>9.9 (2.2)</td>
<td>0.079</td>
<td>-0.11</td>
<td>17.9 (1.4)</td>
<td>0.034</td>
</tr>
<tr>
<td>1-4 completed forms (baseline)</td>
<td>11</td>
<td>9.6 (2.7)</td>
<td>0.744</td>
<td>0.12</td>
<td>18.9 (1.0)</td>
</tr>
<tr>
<td>1-4 completed forms (resultant)</td>
<td>9.9 (2.2)</td>
<td>0.079</td>
<td>-0.11</td>
<td>17.9 (1.4)</td>
<td>0.034</td>
</tr>
<tr>
<td>5-9 completed forms (baseline)</td>
<td>19</td>
<td>9.9 (2.4)</td>
<td>0.774</td>
<td>-0.11</td>
<td>17.9 (1.4)</td>
</tr>
<tr>
<td>5-9 completed forms (resultant)</td>
<td>9.6 (2.8)</td>
<td>0.079</td>
<td>-0.11</td>
<td>18.9 (1.2)</td>
<td>0.034</td>
</tr>
<tr>
<td>≥10 completed forms (baseline)</td>
<td>19</td>
<td>9.9 (2.4)</td>
<td>0.774</td>
<td>-0.11</td>
<td>17.9 (1.4)</td>
</tr>
<tr>
<td>≥10 completed forms (resultant)</td>
<td>9.6 (2.8)</td>
<td>0.079</td>
<td>-0.11</td>
<td>18.9 (1.2)</td>
<td>0.034</td>
</tr>
<tr>
<td>No peer-assessment</td>
<td>31</td>
<td>10.2 (1.7)</td>
<td>0.109</td>
<td>0.40</td>
<td>17.8 (1.2)</td>
</tr>
<tr>
<td>With RQ (baseline)</td>
<td>11.1 (2.5)</td>
<td>0.109</td>
<td>0.40</td>
<td>17.7 (1.9)</td>
<td>0.810</td>
</tr>
<tr>
<td>With RQ (resultant)</td>
<td>10.2 (1.7)</td>
<td>0.109</td>
<td>0.40</td>
<td>17.8 (1.2)</td>
<td>0.810</td>
</tr>
<tr>
<td>≥10 completed forms (resultant)</td>
<td>11.1 (2.5)</td>
<td>0.109</td>
<td>0.40</td>
<td>17.7 (1.9)</td>
<td>0.810</td>
</tr>
</tbody>
</table>
Table 5-18 BDS 5 baseline (2012) and resultant (2013) mean scores, standard deviations (sd), statistical significance of the difference (p) (paired-samples t-test), and the effect size (d) for each of the four scales of the RQ according to participants’ academic year and whether they participated in the peer-assessment exercise.

<table>
<thead>
<tr>
<th>BDS 5 baseline versus resultant Reflection Questionnaire scores</th>
<th>N</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd)</td>
<td>p</td>
<td>d</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>baseline vs. resultant</td>
<td></td>
<td>baseline vs. resultant</td>
</tr>
<tr>
<td>All who completed the RQ (baseline)</td>
<td>86</td>
<td>12.4 (1.4)</td>
<td>0.695</td>
<td>-0.20</td>
<td>17.2 (2.2)</td>
</tr>
<tr>
<td>All who completed the RQ (resultant)</td>
<td>12.1 (1.8)</td>
<td>17.4 (2.4)</td>
<td></td>
<td>17.8 (2.1)</td>
<td></td>
</tr>
<tr>
<td>Peer-assessment</td>
<td></td>
<td>12.7 (2.9)</td>
<td>0.639</td>
<td>-0.20</td>
<td>17.2 (2.8)</td>
</tr>
<tr>
<td>With RQ (baseline)</td>
<td>68</td>
<td>12.2 (2.3)</td>
<td></td>
<td>17.4 (2.3)</td>
<td></td>
</tr>
<tr>
<td>With RQ (resultant)</td>
<td>12.9 (2.6)</td>
<td>17.4 (2.3)</td>
<td></td>
<td>18.2 (2.1)</td>
<td></td>
</tr>
<tr>
<td>1-4 completed forms (baseline)</td>
<td>28</td>
<td>12.3 (2.5)</td>
<td>0.746</td>
<td>0.21</td>
<td>16.8 (3.0)</td>
</tr>
<tr>
<td>1-4 completed forms (resultant)</td>
<td>12.9 (2.6)</td>
<td>16.9 (1.9)</td>
<td></td>
<td>17.5 (1.8)</td>
<td></td>
</tr>
<tr>
<td>5-9 completed forms (baseline)</td>
<td>18</td>
<td>11.7 (1.6)</td>
<td>0.985</td>
<td>0.02</td>
<td>16.5 (3.9)</td>
</tr>
<tr>
<td>5-9 completed forms (resultant)</td>
<td>11.7 (2.2)</td>
<td>18.1 (1.7)</td>
<td></td>
<td>18.1 (1.7)</td>
<td></td>
</tr>
<tr>
<td>≥10 completed forms (baseline)</td>
<td>22</td>
<td>13.6 (2.9)</td>
<td>0.434</td>
<td>-0.65</td>
<td>18.1 (1.2)</td>
</tr>
<tr>
<td>≥10 completed forms (resultant)</td>
<td>11.9 (2.1)</td>
<td>18.3 (1.6)</td>
<td></td>
<td>19.2 (2.6)</td>
<td></td>
</tr>
<tr>
<td>No peer-assessment</td>
<td>18</td>
<td>12.0 (1.5)</td>
<td>0.659</td>
<td>-0.24</td>
<td>16.4 (2.3)</td>
</tr>
<tr>
<td>With RQ (baseline)</td>
<td>11.6 (2.0)</td>
<td>17.1 (2.9)</td>
<td></td>
<td>17.7 (1.2)</td>
<td></td>
</tr>
</tbody>
</table>
Table 5.19 BDS 2 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 31 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd)</td>
<td>p</td>
<td>d</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diff. with no peer-assessment group</td>
<td></td>
<td></td>
<td>diff. with no peer-assessment group</td>
</tr>
<tr>
<td>All who completed the RQ</td>
<td>71</td>
<td>10.6 (2.8)</td>
<td>0.445</td>
<td>-0.16</td>
<td>18.2 (1.6)</td>
</tr>
<tr>
<td>Peer-assessment</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All participating students</td>
<td>40</td>
<td>9.9 (2.4)</td>
<td>0.050</td>
<td>-0.46</td>
<td>18.7 (1.2)</td>
</tr>
<tr>
<td>1-4 completed forms</td>
<td>10</td>
<td>10.4 (1.7)</td>
<td>0.444</td>
<td>-0.28</td>
<td>17.9 (1.1)</td>
</tr>
<tr>
<td>5-9 completed forms</td>
<td>11</td>
<td>9.9 (2.2)</td>
<td>0.187</td>
<td>-0.46</td>
<td>18.9 (1.0)</td>
</tr>
<tr>
<td>≥10 completed forms</td>
<td>19</td>
<td>9.6 (2.8)</td>
<td>0.070</td>
<td>-0.52</td>
<td>18.9 (1.2)</td>
</tr>
<tr>
<td>No peer-assessment</td>
<td>31</td>
<td>11.1 (2.5)</td>
<td>-</td>
<td>-</td>
<td>17.7 (1.9)</td>
</tr>
</tbody>
</table>

BDS 2 Reflection Questionnaire scores 2013
Table 5.20 BDS 5 resultant (2013) mean scores, standard deviations (sd), statistical significance (independent-samples t-test) of the difference (p), and the effect size (d), when compared to the 18 non-participating students (No peer-assessment), for each of the four scales of the RQ according to whether students participated in the peer-assessment exercise and the number of peer-assessment encounters completed.

<table>
<thead>
<tr>
<th>Studied groups</th>
<th>N</th>
<th>Habitual Action</th>
<th>Understanding</th>
<th>Reflection</th>
<th>Critical Reflection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
<td>Mean (sd)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>p</td>
<td>d</td>
<td>p</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td></td>
<td>diff. with no peer-assessment group</td>
<td>diff. with no peer-assessment group</td>
<td>diff. with no peer-assessment group</td>
<td>diff. with no peer-assessment group</td>
</tr>
<tr>
<td>All who completed the RQ</td>
<td>86</td>
<td>12.1 (1.8)</td>
<td>0.659</td>
<td>0.29</td>
<td>17.4 (2.4)</td>
</tr>
<tr>
<td>Peer-assessment</td>
<td></td>
<td>12.2 (2.3)</td>
<td>0.627</td>
<td>0.29</td>
<td>17.4 (2.3)</td>
</tr>
<tr>
<td>All participating students</td>
<td>68</td>
<td>12.9 (2.6)</td>
<td>0.252</td>
<td>0.57</td>
<td>16.9 (1.9)</td>
</tr>
<tr>
<td>1-4 completed forms</td>
<td>28</td>
<td>11.7 (2.2)</td>
<td>0.920</td>
<td>0.06</td>
<td>17.6 (1.8)</td>
</tr>
<tr>
<td>5-9 completed forms</td>
<td>18</td>
<td>11.9 (2.1)</td>
<td>0.823</td>
<td>0.18</td>
<td>18.3 (1.6)</td>
</tr>
<tr>
<td>≥10 completed forms</td>
<td>22</td>
<td>11.6 (2.0)</td>
<td>-</td>
<td>-</td>
<td>17.1 (2.9)</td>
</tr>
<tr>
<td>No peer-assessment</td>
<td></td>
<td>Age</td>
<td>18</td>
<td>11.6 (2.0)</td>
<td>-</td>
</tr>
</tbody>
</table>

BDS 5 Reflection Questionnaire scores 2013
5.6 Discussion

This Chapter reports the utility results of the implementation of the previously piloted (Chapter 3) prospective, formative, structured peer-assessment protocol of undergraduate dental students’ pre-clinical competence and clinical performance, used as a framework for immediate dialogic peer-feedback. Further, and bearing in mind that work-based performance is case-specific (Norcini, 2003a), the protocol was designed to be implemented on a continuous basis so that several encounters could help to produce reliable results. Accordingly, BDS 2 and BDS 5 students practiced, through direct and permanent observation, how to select good evidence (Biggs and Tang, 2011) to make judgements of their peers’ process of simulation exercises (pre-clinical students) and patient care (clinical students), to provide accurate peer-feedback, during the whole academic year. As highlighted previously (Chapter 3, page 129), it is important to remember that students were not asked to judge their peer’s competence or performance “quality”, nor the practice “volume”, which might prove difficult for junior students (Norcini, 2003a), but rather their “increasing ability over time”. However, it was inevitable, but desirable from an educational viewpoint, that the peer-assessment exercise fostered students to self-reflect on the “quality” of their own work (Biggs and Tang, 2011 p. 196).

Consistent with Miller’s concept of work-based assessment (Miller, 1990), in order to avoid an artificial testing situation like in skill-based assessments (Wass, 2011), the design of both peer-DOPS and peer-mCEX forms required a “routine” observation of the actual job activities of students in the cognitive and pre-clinical/clinical skills, communication, professionalism and management domains (Chapter 3, Table 3-1, Table 3-2, Table 3-3, in pages 115-117). Further, it was intended to go beyond the peer-assessment “marking” part of the process, complementing this with its feedback or socially constructed learning component (Cushing et al., 2011), through which skills are developed (Dochy et al., 1999). Accordingly and despite students taking an active role marking the peer-assessment encounters, they were encouraged to use it as a framework for informed dialogic peer-feedback, thus focusing it more on the “learning” rather than on the “assessment” (Liu and Carless, 2006; Driessen and Scheele, 2013).

The dental and medical literature on peer-assessment methods contains many reports (Chapter 1) mainly focused on different aspects of validity, reliability and/or feasibility (Davis, 2002; Archer et al., 2005; Dannefer et al., 2005; Lurie et al., 2006; Evans et al.,
2007; Archer et al., 2008; Davies et al., 2009; Mackillop et al., 2011a; Speyer et al., 2011; Taylor et al., 2013; Ali et al., 2014; Basehore et al., 2014), and students’ perceptions of their effects (Dannefer et al., 2005; Shue et al., 2005; Larsen and Jeppe-Jensen, 2008; Nofziger et al., 2010; Setna et al., 2010; Teich et al., 2014). However, peer-assessment studies focusing on all five variables that define the utility of an assessment method according to van der Vleuten (1996), are lacking. To the authors’ knowledge, this study is the first report to present the comprehensive analysis of a dental undergraduate peer-assessment and peer-feedback protocol’s reliability, validity, feasibility, acceptability, and most usefully, its educational impact.

The absolute G coefficients for pre-clinical peer-DOPS (0.72) and that for clinical peer-DOPS (0.79) were higher than those obtained in the piloting of the peer-assessment protocol (0.62 and 0.67, respectively) (Chapter 3, page 124). However, the current Chapter G coefficients for pre-clinical peer-DOPS, clinical peer-DOPS and clinical peer-mCEX were lower than those previously reported (Moonen-van Loon et al., 2013) for 16 encounters using DOPS (0.88), Mini-CEX (0.89), and MSF (0.88) for the assessment of 953 medical residents.

Notably, all our three formative peer-assessment instruments including pre-clinical peer-DOPS, clinical peer-DOPS and peer-mCEX reached the generally accepted reliable coefficients of ≥0.7 for low-stakes situations (Beard et al., 2011) at 8, 8 and 7 encounters, respectively (Table 5-5), and thus the first hypothesis of this Chapter can be accepted: The implemented structured peer-assessment protocol is reliable.

This is in agreement with the conclusions of Williams et al. (2003) in that a minimum of 7 to 11 judgements of clinical performance are required for reliable findings. Coincidently, our results also demonstrated the need for 10 or more peer-assessment and peer-feedback encounters to positively impact students’ academic performance. This is probably the most important contribution of the present study: work-based, structured and continuous formative peer-assessment and peer-feedback of both pre-clinical and clinical dental undergraduate students has the potential to significantly enhance students’ learning. In the words of Schuwirth (2013), the protocol does not only “make the horse drink” but also “make it want to drink”.

This important finding might be explained by the requirement for students who engaged in the peer-assessment and peer-feedback protocol, to take an active role in the
management of their own learning (Liu and Carless, 2006). Thus, students self-regulate their learning by monitoring their work using their peers’ feedback as a catalyst, and by expressing and articulating what they know or understand (Liu and Carless, 2006). This being consistent with social constructivist conceptions of learning in which social interactions with other learners play a fundamental role in the development of understanding and meaning, bridging the gap of the zone of proximal development through scaffolding (Wood et al., 1976; Vygotsky, 1978) (further explained in Chapter 1, Constructivism in page 31). This is not to deny the possible implication of other learning theories like experiential learning (Kolb, 1984) or reflection in- and reflection on-action (Schön, 1983), explained in Chapter 1 (pages 33-35) and discussed below.

Following Vygotsky’s model (1978), it can be hypothesised that the scaffolding peers provided to each other was, somehow, incremental. As shown in the graphical representation of Vygotsky’s model presented in Figure 1-1 (page 32), the ZPD is the area where the learner can move with help of a more knowledgeable other: the scaffolding. In Chapter’s 1 representation, the ZPD is wide, thinking of the strong scaffold a clinical tutor can provide to the learner, but the edge of the ZPD is a strong border, highlighting a clear limit where the student can go. Now, if we adapt this general model to the results of the present Chapter as interpreted in Figure 5-7, it can be noticed that the ZPD has been divided in smaller areas representing each peer-feedback encounter.

This, considering the “knowledgeable other” is a peer and her or his scaffolding cannot be compared to the tutor’s one. However, the repetitive nature of the current peer-assessment protocol, allows the thinking of the possibility of the represented incremental scaffolding effect. Further, it can also be noticed in this adapted model the dotted edges of each incremental ZPD area in contrast to the strong border of the general model, representing the permeability of the development zone so that students ‘can’ go a bit further. Finally, the model has been drawn containing ten growing ZPD areas to represent the need for the same number of peer-encounters to obtain a measurable educational impact.
From a different perspective and based on Nicol and Macfarlane-Dick (2006) model of self-regulated learning, students in our study might have taken a proactive (rather than reactive) role in developing their objectivity in the continuous exercise of observing and providing feedback on their peers’ work, which could have then been transferred to their own pre-clinical or clinical work, enhancing their learning (Falchikov, 1995), and potentially improving performance in high-stakes assessments (Liu and Carless, 2006).

This enhancement of student learning as a result of peer-assessment and face-to-face immediate dialogic peer-feedback on each other’s work, has been reported to occur by means of diplomatic criticism, analysis and reflection (Falchikov, 1995; Dochy et al., 1999).

The same finding may explain the statistically significant increase and positive effect size in reflective skills between those students who completed a higher number of peer-assessment and peer-feedback encounters as compared to those who did none, which again allows for the sixth hypothesis of the Chapter to be accepted: The peer-assessment

---

*Figure 5.4 Graphical interpretation of Vygotsky ZAD (Zone of Actual Development) and ZPD (Zone of Proximal Development) (Vygotsky, 1978) according the results of the present chapter.*
exercise stimulates students’ reflective skills.

Consequently, keeping the focus on the learning effects of the protocol, the impact of the current peer-assessment and peer-feedback intervention was especially important for the former groups (those with $\geq 10$ encounters). Accordingly, the group of 19 BDS 2 students who completed 10 or more peer encounters demonstrated both a significantly better performance and medium effect sizes in their high-stakes MCQs ($p=0.047$, $d=0.49$), Clinical Skills Examination viva ($p=0.050$, $d=0.48$) as well as practical ($p=0.21$, $d=0.56$) (Table 5-15), but also a statistically significant increase and large effect size in their Understanding ($p=0.011$, $d=0.71$) and Reflection ($p=0.0007$, $d=1.12$) skills (Table 5-19), as compared to those who did not participate in the peer-assessment exercise.

Indeed, MCQs were aimed at testing the degree of understanding of a concept or topic from different aspects of the course by knowledge recall, association and classification of concepts. Further, the Clinical Skills Examination practical judged the students’ clinical competence in assessing and restoring a tooth in the phantom head and was thus quite similar to the continuous peer-assessment and peer-feedback protocol domains. The viva required students to answer clinical /materials /instrument questions related to a given clinical problem thus necessitating students to not only recall and understand, but also to apply knowledge in order to solve the clinical problem. With a different aim, written SAQs and SNQs are designed to assess knowledge recall which contrasts with the more application of knowledge focus of the pre-clinical peer-DOPS form. This might explain why the peer-assessment exercise did not produce any measurable effect in any of the studied groups. Further, going back to the preceding Chapter 4 (Table 4-7 in page 145), Paper 1a, consisting of short answer questions and short note questions for BDS 1 to 3, similarly did not correlate to any of the RQ constructs.

Interestingly, the group of 11 BDS 2 students who completed between 5 and 9 peer encounters embodied a large effect size of their Understanding ($d=0.77$) and Reflection skills ($d=0.52$) in a year time. However, this was not reflected in a better examination performance, perhaps suggesting the need for a longer period of continuous reflection.

In a similar trend, those 22 BDS 5 students who performed $\geq 10$ peer-DOPS and/or peer-mCEX exercises exhibited a significantly better performance in their high-stakes clinical scenario Essays, MCQs and Clinical Reasoning Examinations, when compared
to the rest of their class. Furthermore, they also showed a significant increase in their Critical Reflection skills, which was not seen in any other group. This might indicate students reached the deep learning qualitative, “relational” and then “extended abstract”, phase of the SOLO taxonomy (Biggs and Tang, 2011 pp. 86-91), which has been described as a useful complement to the Kember et al. (2008) RQ when teaching and assessing understanding, construction and interpretation.

Once more, as observed in Chapter 4 (Table 4-7 in page 145), where the OSCE results correlated to both RQ Habitual Action and Understanding scales, this chapter’s imperceptible impact of the peer-assessment protocol on OSCE’s scores, was unforeseen. However, the answer to this might be found precisely in its positive relationship with Habitual Action and Understanding. As shown in Table 5-18, the peer-assessment and peer-feedback protocol only showed a significant difference and a large effect size ($p=0.003$, $d=1.04$) in the Critical Reflection scale for those BDS 5 students with 10 or more encounters. It would then be peculiar to find an increase in OSCE’s scores when the evidence from the previous chapter showed they are only related to Habitual Action and Understanding. Furthermore, the same 22 students who peer-assessed each other 10 or more times and demonstrated a large Critical Reflection effect size, also showed a large negative size effect of $d=0.65$ in their Habitual Action scores. This could be interpreted as a move from automated responses to becoming aware of why they perceive, think, feel or act as they do (Kember et al., 2008).

BDS 2 Clinical Skills Examination viva section and the BDS 5 Essay and Clinical Reasoning Examinations have in common the requirement that students solve problems or cases. Thus, as presented by Mezirow (1991 pp. 99-117), this leads to higher order thinking, as the central dynamic involved in problem solving is reflection.

However, there still remains the question why BDS 2 students did not experience an improvement or higher effect size in their Critical Reflection skills, as did BDS 5 students. The answer to this might be found in the different environments both groups of students conducted the peer-assessment protocol, or rather what happened next. Following the trained protocol, and circumscribing it to Kolb’s four-stages experiential learning cycle by which knowledge is created through the transformation of experience (Kolb, 1984 p. 41) (Chapter 1, Experiential learning in page 33), both groups of participating students observed/performed the pre-clinical or clinical task (Kolb’s first stage: concrete experience), peer-assessed and debriefed the experience during the
course of the peer-feedback (Kolb’s second stage: reflective observation), followed by an assimilation and distilling of the feedback information through self-assessment/reflection to draw new implications for actions (Kolb’s third stage: abstract conceptualization). Finally, these new implications must be actively experienced for the learning to be effective (Kolb’s fourth stage: active experimentation). However, in our protocol this stage of the cycle did not happen immediately after the abstract conceptualization, if it happened at all, as the third stage was performed at the end of each pre-clinical or clinical session. Hence, according to Kolb, conceptualized (3rd stage) but not experimented (4th stage) new implications/ changes/ frames may be missed (Kolb, 1984).

Returning to our protocol, BDS 5 students spent 68% of their time working with patients in the clinic (Chapter 2, Figure 2-4 in page 99), meaning their conceptualizations (3rd stage) could be experimented (4th stage) right after and in a daily basis. In sharp contrast, BDS 2 students’ conceptualizations could only be experimented one week later, as they attended the pre-clinical skills laboratory on weekly bases. Further, according to Jarvis’s (1987) revised Kolb’s cycle, and due to this missing or delayed experimentation 4th stage of the latter students, it might be argued that they involuntarily adopted a “non-reflective learning” practice instead of the intended reflective learning. The author’s model describes “non-reflective learning” as an acquisition of physical (manual) skills without reflection, which is the most common form of learning that occur in everyday life, but at the same time one that does not result in innovation (Jarvis, 1987). This contention might be supported by the present Chapter results in which participating BDS 2 students demonstrated the most significant results ($p=0.021$) and highest effect size ($d=0.56$) in their Clinical Skills Examination practical, which judges the students’ practical skills to restore a tooth in the phantom head.

In a different but complementary theoretical perspective, BDS 5 participating students might have reached the critical reflection level of thinking, as they were working and practicing the peer-assessment protocol in the very relevant and problem-centred environment of the patients’ clinic. Thus according to Dewey’s theory of reflective thought and action (Dewey, 1909; Dewey, 1938), they might have been most motivated to reflect to resolve the conflict when facing the “inadequacy” or “state of uncertainty” (that is failures and difficulties) of clinical results (Dennick, 2012). At a higher level, if we think of this “inadequacy” as unusual or unexpected problems, surprises or even conflicts while treating a patient, they might have triggered the Schön (1983; 1987)
Chapter 5 Implementing a Structured Protocol for Dental Students’ Prospective Peer-Assessment and Peer-Feedback

concepts of reflection-in- and reflection-on-action (Chapter 1, Reflective practice learning, in page 35), where participants examined the what, the how and the alternatives of the patient encounter. However, as he further states, in the same situation of reflection-in- and reflection-on-action of an ill-defined problem but adding the opportunity to share and process thoughts from the experienced problem with others (peers and facilitators), it is possible to reach the higher critical reflection level of thinking, questioning the why and the beliefs either helping or hindering the work at the encounter. Academically important, this is highly unlikely to happen without the pressure of inadequacy and facilitation through others (Miettinen, 2000; Stocker et al., 2014).

There is no published data available to compare the educational impact of the current dental students’ peer-assessment and peer-feedback protocol. However, a meta-analysis on 41 studies on medical clinical performance (Veloski et al., 2006), reported that 32 of the reports demonstrated a positive impact of feedback, but the variation in outcome variables impeded any systematic analysis of effect sizes. Somehow further, the meta-analysis of feedback interventions by Kluger and DeNisi (1996) discussed in Chapter 1 (Feedback that works, in page 70) found the averaged effect size of feedback interventions across four groups of variables (cues, task characteristics, situational, and methodological) to be $d=0.38$, though peer-feedback or peer-tutoring was not within the moderators.

However, the huge study by Hattie (2012), made it possible to relate the present chapter’s effect size results to his “list of influences on achievements”. Thus, in his study the influence of feedback in classroom had an effect size of $d=0.75$ (especially when it is task focused and not to the self) and is ranked number 10 in a list of 150 influences, whereas “peer tutoring” has an effect size of $d=0.55$ and is tiered number 34. As it can be noticed, both interventions have an effect size well above the $d=0.40$ “gold standard” (Hattie and Timperley, 2007). Moreover, a just published meta-analysis of 31 studies on feedback for simulation-based procedural skills training which included six on dental students –though none on peer-feedback- found a positive effect size of $d=0.74$ in favour of feedback (Hatala et al., 2014).

Accordingly, BDS 2 participating students who peer-assessed and provided peer-feedback to each other 10 or more times, exhibited an effect size of $d=0.80$ for the average of all their high-stakes examinations, as compared to those who did not
participate. Similarly, those BDS 5 participating students with 10 or more peer encounters revealed an effect size of $d=0.74$ for the average of all their Finals. Consequently, the fifth hypothesis of the Chapter is accepted: The implemented structured peer-assessment protocol has a positive educational impact.

Returning to the psychometrics of the peer-assessment protocol, notwithstanding that previous studies have reported the validity of the conventional formats of both Mini-CEX (Norcini et al., 2003; Al Ansari et al., 2013) and DOPS instruments (Wilkinson et al., 2008; Delfino et al., 2013), the new peer-DOPS and peer-mCEX rating scale and modified domains required a new validation process. Thus, content validity was ensured by the fact that all three forms were suitable for 99% of the encounters allowing a broad range of pre-clinical and clinical situations and procedures (Williams et al., 2003) (Table 5-2 and Table 5-3).

Criterion validity, as a correlation between peer-assessment scores and high-stakes end-of-year examinations, was statistically significant for both peer-DOPS forms but not for the peer-mCEX, despite BDS 5 peer-DOPS and peer-mCEX scores were very similar. This inconsistency might be explained by the low number of 77 peer-mCEX completed encounters compared to 576 peer-DOPS encounters. It should be borne in mind that undergraduate dental education is based on strong procedural practice, initially in simulators (phantom heads) and subsequently on real patients (KCLDI teaching methodologies by course year in page 99). Clinical students perform a patient’s clinical examination/assessment and implement a treatment plan during the successive clinical sessions. This explains the big difference between peer-DOPS and peer-mCEX numbers. From this perspective, it is probably correct to think that the inclusion of the peer-mCEX for undergraduate BDS 5 students was not well thought out.

Regarding construct validity, both groups of students using peer-DOPS and peer-mCEX forms demonstrated a consistent ability to detect significant differences in their peers’ performance in the diversely assessed domains (Table 5-7, Table 5-8, Table 5-9), despite Sluijsmans et al. (2002) hypothesis that novices have a lower capability to assess certain domains. However, and as the same authors ascertained, training students in assessment skills positively affects their performance, which is exactly what our protocol did. Indeed, by using written/video examples and role-playing, all participating students in the present study received an intense training and familiarisation session on observation, peer-assessment, peer-feedback, action plan and completion of the peer-
DOPS and peer-mCEX instruments. As mentioned above, every peer-assessment session was monitored by the same researcher, who performed the training in a non-intrusive manner and thus any needed reinforcement was immediately done. Further, the educationally referenced scale containing a graphical and written anchor with the desired “increasing ability over time” (Chapter 3, Figure 3-6), might have had a positive impact in helping students to develop their assessment skills and explain their discrimination judgement. This possible indication of alignment between the scale and the students’ expertise (Crossley et al., 2011) might explain why participating students agreed or strongly agreed on the validity of the peer-assessment protocol in contrast with the study by Bennett et al. (2012) in which undergraduate medical students peer-assessed each other using the standard Mini-CEX norm-referenced scale (Below expectation, Borderline, Meets expectations and Above expectations).

Moreover, and possibly due to the same aforementioned reasons, students also revealed their capability to perceive significant changes in performance over time as early as the initial sessions. This was particularly interesting as the pattern was dissimilar between pre-clinical and clinical peer-DOPS and peer-mCEX scores (Figure 5-3). Thus, at the beginning of the academic year (September), BDS 5 peer-assessment scores were higher for patients’ clinical examinations and treatment planning assessed using peer-mCEX than those given for clinical procedures by peer-DOPS. Later, both scores showed an upward gradient, though peer-DOPS’ scores were steeper reaching those of peer-mCEX by November. This might indicate students felt initially more comfortable assessing patients than performing procedures. Further, by December/January both scores had reached a plateau (means 5 to 6) which continued as such until the end of the study. Surprisingly, these high and stable scores were reached during the same period as the students gave their Dental Foundation Training interviews which could be interpreted as they considered themselves ready as safe beginners.

In contrast, and likewise the same peer-assessment protocol preceding pilot study (Chapter 3, Figure 3-7 in page 126), BDS 2 students only started to notice peer improvement at the fifth session, while during the first four encounters scores dropped significantly. As mentioned before, this difference might be explained by an initial calibration process (Hauser and Bowen, 2009), the needed time to gain experience as evaluators (Karl et al., 2011), and becoming at ease with making their work public (Liu and Carless, 2006), or even adjusting to the new pre-clinical learning environment (Schoenrock-Adema et al., 2007).
Altogether, both undergraduate groups in the present study showed a similar capacity to perceive an increasing ability over time as that reported by Davies et al. (2009) in the already graduated foundation medical trainees’ peer-assessment (using the mini-PAT), and that of Playford et al. (2013) and Prescott-Clements et al. (2011) in which staff trainers continuously assessed undergraduate medical students and postgraduate dental trainees, respectively, with similar clinical assessment forms. Consequently, and in light of the above discussed validity components of the protocol, the second hypothesis of the Chapter can be accepted: The implemented structured peer-assessment protocol is valid.

The overwhelming students’ perception of the protocol’s acceptability and fairness, expressed through the anonymous end-of-study questionnaire, was unexpected and allows to admit the fourth hypothesis of the Chapter: The implemented structured peer-assessment protocol is acceptable. In addition, no student stopped participating in the study or asked for the return of their completed peer-DOPS and peer-mCEX forms, having the freedom to do so as stated in their signed consent forms. As stated by Biggs and Tang (2011), this high acceptability could be interpreted as students “became accustomed to being observed by one another when they gave and received peer-feedback”.

The explanation to this might be found in the other answers to the same anonymous questionnaire. Thus, students might have found the peer-assessment exercise acceptable as they largely scored it to be “easy and straightforward to use”, and that it helped them “to identify learning needs and to improve their performance”. Another possible reason can be found in the study by Shue et al. (2005), who concluded that “most students are willing to participate in peer-assessment as long as their preferences are taken into consideration”. As a matter of fact, during the whole study period, students were always heard when they wanted to ask or give a comment on how to improve the protocol.

Besides the positive psychometric properties described above, there are also some practical reasons to implement a peer-assessment and peer-feedback protocol. Thus, in the presence of resource constraints and staff problems to provide sufficient feedback (Liu and Carless, 2006), peer-assessment and feedback can help to overcome the frequently reported lack of immediate feedback in workplace-based assessment as well as difficulties in finding a staff assessor (Quantrill and Tun, 2012). Furthermore, previous studies have reported a box ticking attitude when members of staff complete clinical assessment forms (Sabey and Harris, 2011), which was not observed or reported
Given these points and the present study’s positive impact on students’ academic results, supported by all other utility variables, peer-assessment and peer-feedback have significant potential in the current dental education setting. A frequently mentioned innovation in pre-clinical and clinical dental education is that of students working in pairs or teams mimicking the real working environment (Coomarasamy and Khan, 2004; Haden et al., 2010) and assisting others to learn. Thus, the implementation of a peer-assessment and peer-feedback in such naturalistic settings would only mean structuring and organising a protocol without any changes in curricular content, which otherwise would mean losing an opportunity to foster students’ academic and reflective skills. In other words, this would mean moving the serendipitous interactions between students, which would inevitably happen, from the uncontrolled “hidden” to the stated “formal” curriculum (Hafferty and Franks, 1994).

Future research may address some variables that this study did not contemplate. Peer-assessment and peer-feedback information could be used for an early exposure of those students needing tutor remediation. Further, as suggested by some students in the present study (Table 5-12), allocating specific time slots at the end of the session to allow peer-assessment and peer-feedback, and moving into electronic peer-DOPS and peer-mCEX forms, could attract more students or even allow them to complete more encounters. The latter student’s suggestion might also help in easing the excessive amount of time and effort required to implement and monitor the current protocol. It would also be helpful to know whether peer-assessment or reflective skills scores have any predictive value in identifying the pre-clinical student who will do well in the clinic (Fugill, 2013) and in her/his future professional life.

5.7 Conclusions

The current study protocol of continuous, formative and structured pre-clinical and clinical peer-assessment as a framework for immediate dialogic peer-feedback of undergraduate dental students is not only reliable, valid, feasible and acceptable but it also has the potential to make a significant educational impact, provided it is practiced during ten or more peer encounters. In this instance, development of students’ academic and reflective skills can be expected.
These results suggest that dental undergraduate students can be given more responsibility to take an active role in the low-stakes formative phase of learning by moving from being passive students to active trainers, as in the present peer-assessment and peer-feedback protocol, especially if our aim is to teach them to self-regulate and control their learning and so be prepared to manage their own education throughout life.

Thus, in this paradigm of enabling students’ self-learning in parallel to a traditional delivered instruction, the next Chapter will further analyse the peer-feedback narratives students provided to each other to try to elucidate any clues that may help to better understand the present Chapter results and further develop the peer-assessment protocol.
Chapter 6

Analysis of Dental Students’ Peer-feedback Narratives from a Prospective Peer-Assessment Protocol

Figure 6-1 Flowchart of the study’s research chapters. The sixth chapter is highlighted to help in providing an overall view.

6.1 Introduction

The move towards assessment for learning rather than assessment of learning has been described as “nothing short of a revolution in the conceptual framework of assessment” (Schuwirth and van der Vleuten, 2011b). In this approach assessment aims to drive the learning experience and so meaningful qualitative feedback is at the heart of the process (Assessment purposes in page 43) (Bok et al., 2013).

However, there are many factors influencing the interpretation and uptake of such
feedback (Eva and Regehr, 2011). These factors include its positive or negative signs (Kluger and Van Dijk, 2010), students motivations (ten Cate et al., 2004; Playford et al., 2013), its credibility and nature of the task (Watling, 2014), evidence and specificity (Delva et al., 2013), delivery moment and frequency (Hatala et al., 2014), feasibility (Smither et al., 2005), receptor focus (Kluger and Van Dijk, 2010), task perception (Pelgrim et al., 2014) and emotional impact (Urquhart et al., 2014).

The complex mixture of these and other factors can have a paradoxical effect (Eva and Regehr, 2011), reinforcing the notion that there is no simple recipe for the delivery of feedback (Boud and Molloy, 2013; McKenzie, 2013). Further, the interplay of aspects might explain that all feedback is not good feedback as it can both help or hinder learning (Kluger and DeNisi, 1996) (Chapter 1, Feedback in page 62).

Against this background and our findings in Chapter 5, we wanted to determine how our peer-assessment and face-to-face immediate peer-feedback protocol could have influenced students’ skills. We therefore investigated the way students constructed, focused and based their feedback narratives.

### 6.2 Aim

The aim of this Chapter was to quantitatively and qualitatively analyse students’ written peer-feedback narratives from the peer-assessment exercise, as well the reasons students decided to take part in, or not take part in the peer-assessment protocol.

### 6.3 Hypotheses

i. Pre-clinical and clinical dental students participating in the peer-assessment protocol provide only general and positive peer-feedback in the clinical domain.

ii. Participating students decided to take part in the peer-assessment protocol as a result of intrinsic and extrinsic motivations.

iii. Students who decided not to take part in the peer-assessment protocol did so as a result of intrinsic and extrinsic motivations.
6.4 Materials and Methods

6.4.1 Quantitative coding of peer-feedback narratives

All 40 BDS 2 and 68 BDS 5 students (mean age=23.5, sd=2.6, 70 females, 38 males) who participated in the year-long, structured peer-assessment and peer-feedback protocol (Chapter 5), completed a total of 1169 peer-DOPS (N=516 pre-clinical and N=576 clinical) and peer-mCEX (N=77) forms. Of these, 1104 (94.4%) included a written feedback comment. These comments were transcribed to a word processor file, their words counted and the feedback narrative content coded by the same researcher (JT), according to the four GDC “Preparing for Practice” (2012a) domains (Clinical, Communication, Professionalism, Management & Leadership).

The peer-feedback narratives’ sign was categorised as positive, when providing information about success; negative, when delivering information about failures; or mixed (Kluger and DeNisi, 1996). The narratives’ specificity was also classified as task specific, when the procedure peers were performing was clear and/or gave a detailed comment on what and/or why something went well or otherwise; or general, when the procedure could not be identified and/or there was no reference to what and/or why something went well or badly. These factors were then analysed separately for BDS 2 and BDS 5 students using descriptive statistics.

6.4.2 Qualitative thematic analysis of peer-feedback narratives

To enable this analysis, the 1104 written feedback comment from BDS 2 and BDS 5 students were grouped into themes (by the same researcher), using an “open coding” interpretative process to break down the data analytically (Corbin and Strauss, 1990) and ensuring all manifestations of each theme had been accounted for (Pope and Mays, 2013). Feedback comments were then compared and those conceptually similar were labelled and grouped together to form themes. These emerging themes were categorised according to the four GDC domains (General Dental Council, 2012a).

6.4.3 Focus groups on the reasons peer-assessment participating and non-participating students decided to take part or not in the protocol

Of the 154 BDS 2 and 155 BDS 5 students invited to participate in the peer-assessment study, 40 (26% of the class) and 68 (44% of the class), respectively, agreed to take part in the protocol. For future implementation purposes, the reasons why students decided
to participate or not were investigated.

During April 2013 (before the peer-assessment and peer-feedback study ended), all BDS 2 and BDS 5 students were invited for a snack lunch, independently of their participation in the study, making it clear that the purpose was to gain in-depth understanding of the reasons that guided them to take part or not in the peer-assessment protocol. These 40-minute meetings were organised and facilitated by the same researcher involved in the study and were performed on two consecutive weeks on Tuesdays and Fridays for BDS 5 (completing 6 meetings) and Thursdays for BDS 2 (completing 2 meetings). Particular care was taken to ensure that each group contained students who participated in the peer-assessment protocol as well as those who did not. Meetings were not recorded or videotaped and the facilitator took written notes of the main themes appearing. Students signed a consent form which stated that their comments would be kept anonymous.

To stimulate the discussion, the facilitator opened each session by inquiring who of the students present did take part in the peer-assessment and peer-feedback protocol and who did not, and why?

As with the feedback narrative analysis, notes from the 8 focus group discussions were compared, labelled and grouped together in theme categories (Corbin and Strauss, 1990).

6.5 Results

6.5.1 Quantitative coding of peer-feedback narratives

Peer-feedback narrative mean number of words was 19.3 (range=2-44) for BDS 2 and 25.2 (range=6-56) for BDS 5. The most common BDS 2 content peer-feedback narratives related to the GDC Clinical domain (81%) and related to practical clinical skills (58%) and clinical knowledge (23%). In contrast, the most common BDS 5 content narrative related to the Management & Leadership domain (36%), closely followed by Communication (32%) (Table 6-1). It should also be noted that whilst most feedback comments related to a single domain, some did mention more than one domain competency, some examples of which are cited below.
Table 6-1 Proportional quantitative coding of BDS 2 and BDS 5 peer-feedback narratives according to their content, sign and specificity.

<table>
<thead>
<tr>
<th>Peer-feedback Narrative Code</th>
<th>BDS 2</th>
<th>BDS 5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content</strong> <em>(GDC domain)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinical Practical</td>
<td>58%</td>
<td>12%</td>
</tr>
<tr>
<td>Clinical Knowledge</td>
<td>23%</td>
<td>0%</td>
</tr>
<tr>
<td>Communication</td>
<td>3%</td>
<td>32%</td>
</tr>
<tr>
<td>Professionalism</td>
<td>1%</td>
<td>12%</td>
</tr>
<tr>
<td>Management &amp; Leadership</td>
<td>10%</td>
<td>36%</td>
</tr>
<tr>
<td>Overarching competencies</td>
<td>5%</td>
<td>8%</td>
</tr>
<tr>
<td><strong>Sign</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>91%</td>
<td>65%</td>
</tr>
<tr>
<td>Negative</td>
<td>2%</td>
<td>13%</td>
</tr>
<tr>
<td>Mixed</td>
<td>7%</td>
<td>22%</td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task specific</td>
<td>77%</td>
<td>74%</td>
</tr>
<tr>
<td>General</td>
<td>23%</td>
<td>26%</td>
</tr>
</tbody>
</table>

**BDS 2:**

“Successfully placed on clamp, however need to remember to floss between each tooth and work a lot faster”

“Well time managed and very good manual dexterity and posture. Need to study more on tooth morphology & mixing materials”

“You really need to work on posture and go for harder challenges as it is the best way to learn. I also need it so let’s challenge each other every Thursday”

“Practice more excavation as we just did to remove caries so you are sure in distinguishing between soft infected dentine and slightly harder affected dentine using an excavator”

“Continue practicing caries removal to enhance both time management and technique by improving assessment of depth of caries lesions as agreed on last session”

“Try and be quicker when carving and getting the right anatomical structure before the GIC sets”

“Very good use of time for the fissure sealants using all available materials, but should remember to use rubber dam”

**BDS 5:**

“Excellent work but need to be more confident”

“Don’t be afraid to ask for help and remember to vaseline patient’s lips before placing disclosing tablet”
“Very well use of electrosurgery for the first time. However, don't forget time management”

“Difficult subgingival filling that entered into the pulp and you did well to notice this well in advance and asked for help. Remember to notice a pulp is necrotic from smell and do more tooth vitality tests before treatment”

“Try to get better in decision making and improve your confidence. We worked as a team but you will soon be alone”

“Remember to cover ultrasonic scales tip when not in use”

“Good communication with patient throughout the treatment and kept calm when struggling. You knew what you wanted to do, but as the treatment went difficult you should have asked for help at an early stage, don’t compromise”

In relation to the ‘sign’ of peer-feedback narratives, 91% of those from BDS 2 students and 65% from BDS 5 students were categorised as positive, while 7% and 22%, respectively, being mixed. Some selected citations of the three sign categories (Table 6-1) are reproduced below:

Positive:

“Good foundation of understanding about the procedures. Suggest improving skills with hand pieces - positioning, time keeping and control of materials”

“You can do more than you think. Just be more confident in your own abilities”

Negative:

“At this stage you should already be able to work in a systematical manner”

“You lack management as should have booked double slot for this difficult RCT”

Mixed:

“Colour of the tooth nice, cusp pattern close but not morphologically correct”

“Excellent communication with patient, but if you wanted to do vitality test, make sure it is done before patient is anaesthetised. Think ahead!”

Similarly, most of BDS 2 and BDS 5 feedback comments were classified as task specific (77% and 74%, respectively) (Table 6-1). A few examples of these and other general comments were:
Task specific:

“You exhibited great competence using the incremental technique when restoring deep cavities. Continue working on and developing ability to finish composite restoration following the natural UL6 tooth morphology particularly in areas of cuspal build ups”

“After guidance was given from the tutor, you took them on board to extract the UR7 intact. However, you need to develop more clinical skills in XLA of a non-mobile tooth”

General:

“Improve knowledge of dental materials”

“Need to improve management”

6.5.2 Qualitative thematic analysis of peer-feedback narratives

All BDS 2 students’ feedback narratives were grouped into 14 themes representing all GDC domains (General Dental Council, 2012a). There were 9 themes related to the Clinical domain, three to Management & Leadership, and one each to Communication and Professionalism (Table 6-2).

Table 6-2 Themes identified in BDS 2 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).

<table>
<thead>
<tr>
<th>BDS 2 Themes</th>
<th>GDC domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caries removal</td>
<td>Clinical</td>
</tr>
<tr>
<td>Cavity preparation</td>
<td>Clinical</td>
</tr>
<tr>
<td>Communicate better</td>
<td>Communication</td>
</tr>
<tr>
<td>Correct procedure steps</td>
<td>Clinical</td>
</tr>
<tr>
<td>Dental material properties</td>
<td>Clinical</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Management &amp; Leadership</td>
</tr>
<tr>
<td>Harder challenges</td>
<td>Professionalism</td>
</tr>
<tr>
<td>Infection control</td>
<td>Clinical</td>
</tr>
<tr>
<td>Manual dexterity</td>
<td>Clinical</td>
</tr>
<tr>
<td>Practice more</td>
<td>Clinical</td>
</tr>
<tr>
<td>Restoration anatomy</td>
<td>Clinical</td>
</tr>
<tr>
<td>Time management</td>
<td>Management &amp; Leadership</td>
</tr>
<tr>
<td>Working speed</td>
<td>Management &amp; Leadership</td>
</tr>
<tr>
<td>Working position</td>
<td>Clinical</td>
</tr>
</tbody>
</table>
In contrast, BDS 5 students’ feedback narratives were grouped into 24 themes representing all GDC (General Dental Council, 2012a) domains as well (Table 6-3). More specifically, sixteen themes were associated with the Clinical domain, six to Professionalism, two each to Communication and Management & Leadership, and one Overarching outcome.

Table 6-3 Themes identified in BDS 5 peer-feedback narratives presented in alphabetical order (not frequency) and their corresponding GDC domain (General Dental Council, 2012a).

<table>
<thead>
<tr>
<th>BDS 5 Themes</th>
<th>GDC domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anticipate events</td>
<td>Professionalism</td>
</tr>
<tr>
<td>Ask for help</td>
<td>Clinical / Professionalism</td>
</tr>
<tr>
<td>Clinical skills</td>
<td>Clinical</td>
</tr>
<tr>
<td>Communicating with patient</td>
<td>Communication</td>
</tr>
<tr>
<td>Decision making</td>
<td>Clinical</td>
</tr>
<tr>
<td>Diagnostic skills</td>
<td>Clinical</td>
</tr>
<tr>
<td>Compromising patient</td>
<td>Professionalism</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Management &amp; Leadership</td>
</tr>
<tr>
<td>Informed consent</td>
<td>Clinical / Communication</td>
</tr>
<tr>
<td>Infection Control</td>
<td>Clinical</td>
</tr>
<tr>
<td>Knowing your limits</td>
<td>Clinical / Professionalism</td>
</tr>
<tr>
<td>Laboratory work evaluation</td>
<td>Clinical</td>
</tr>
<tr>
<td>Local anaesthesia</td>
<td>Clinical</td>
</tr>
<tr>
<td>Managing complications</td>
<td>Clinical</td>
</tr>
<tr>
<td>Managing patient’s expectations</td>
<td>Clinical</td>
</tr>
<tr>
<td>Overall view</td>
<td>Overarching outcome</td>
</tr>
<tr>
<td>Organisation</td>
<td>Clinical</td>
</tr>
<tr>
<td>Patient anxiety</td>
<td>Clinical</td>
</tr>
<tr>
<td>Professional attitude</td>
<td>Professionalism</td>
</tr>
<tr>
<td>Self confidence</td>
<td>Clinical / Professionalism</td>
</tr>
<tr>
<td>Time management</td>
<td>Management &amp; Leadership</td>
</tr>
<tr>
<td>Understanding procedure</td>
<td>Clinical</td>
</tr>
<tr>
<td>Vitality test</td>
<td>Clinical</td>
</tr>
<tr>
<td>Working speed</td>
<td>Clinical</td>
</tr>
</tbody>
</table>

6.5.3 Focus groups on the reasons peer-assessment participating and non-participating students’ decided to take part or not in the protocol

A total number of 36 (9 BDS 2 and 27 BDS 5) students attended the focus group meetings. Of these, 25 (6 BDS 2 and 19 BDS 5) followed the peer-assessment protocol and 11 (3 BDS 2 and 8 BDS 5) did not participate. The atmosphere in the meetings was
always very positive and participants were honestly willing to give their views.

Those 25 students who participated in the peer-assessment and peer-feedback protocol gave a wide range of reasons to engage (Table 6-4). From these, there were three themes which emerged in all 8 meetings: “enjoy teaching”, “the need for more feedback”, and “trust my peers’ assessment”.

Table 6-4 Themes identified during the focus groups as the reasons why students who followed the peer-assessment protocol decide to take part in it. Recurrent themes are shown in bold.

<table>
<thead>
<tr>
<th>Themes from participating students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoy teaching</td>
</tr>
<tr>
<td>Induction tutorial was convincing</td>
</tr>
<tr>
<td>Informal peer-tutoring was done anyway</td>
</tr>
<tr>
<td>Study seemed innovative</td>
</tr>
<tr>
<td>The Kindle Fire HD incentive</td>
</tr>
<tr>
<td><strong>The need for more feedback</strong></td>
</tr>
<tr>
<td>The research certificate for portfolio</td>
</tr>
<tr>
<td>To help the PhD student in charge</td>
</tr>
<tr>
<td><strong>Trust my peers’ assessment</strong></td>
</tr>
<tr>
<td>Wanted to improve and develop clinical performance</td>
</tr>
</tbody>
</table>

By contrast, the 11 students who did not participate in the peer-assessment study but attended the snack lunch focus groups provided a limited number of reasons why they had not participated in the protocol. However, a common theme was in relation to bad prior formal experience of work-based assessment and feedback (though not necessarily coming from peers), particularly from those students with a previous university degree (Table 6-5).

Table 6-5 Themes identified during the focus groups as the reasons why students decided not to take part in the peer-assessment study. Recurrent theme is shown in bold.

<table>
<thead>
<tr>
<th>Themes from not participating students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bad previous experience with WPBA and/or feedback</strong></td>
</tr>
<tr>
<td>Didn’t think it was helpful</td>
</tr>
<tr>
<td>Didn’t want to take the responsibility</td>
</tr>
<tr>
<td>Felt unable to assess peers</td>
</tr>
<tr>
<td>Have an IPad</td>
</tr>
</tbody>
</table>
6.6 Discussion

This Chapter reports the quantitative and qualitative analysis of students’ written peer-feedback narratives from all those who exercised the peer-assessment protocol. It has also provided some insight into the reasons students decided to take part in, or not take part in the study.

The quantitative coding of both BDS 2 and BDS 5 peer-feedback narratives included their content, sign and specificity. The first analysis (content) of the results showed a notable difference between pre-clinical and clinical students’ narratives. In a way, this might be expected as the environments where each group of students performed their work is totally different: BDS 2 in the pre-clinical skills laboratory with its manual dexterity challenges, and BDS 5 in “the real” world clinic with time and patient stresses. However, this may not fully explain the difference. The overwhelming 81% of Clinical domain in BDS 2 narratives, contrasts with the much lower 12% in BDS 5 ones and could be explained by the level of training of both groups of students. Hence, pre-clinical students are at the beginning of the clinical skills ladder whilst senior students are much further up. Additionally, the fact that BDS 5 students had to work with a patient clearly influenced the content of the provided feedback, focusing it more on other needed skills like Communication, Management & Leadership and Professionalism. This students’ ability to provide peer-feedback not only in the clinical domain already allows for the first hypothesis to be rejected: Pre-clinical and clinical dental students participating in the peer-assessment protocol provide only general and positive peer-feedback in the clinical domain.

Unfortunately, there is a notable absence of evidence on the “content” of feedback studies (Hatala et al., 2014) with which to compare our results. However, the evidence from the “Guidance Hypothesis” from the motor learning literature (Schmidt and Lee, 2011 as cited by Hatala et al. 2014), shows that terminal feedback - given at the end of the practice - is superior to concurrent feedback – given during the practice - for long term skill retention, as the latter may lead to an over-reliance on feedback (Hatala et al., 2014). These findings support our peer-assessment and peer-feedback protocol as the feedback was provided at the end of the pre-clinical or clinical session. This was encouraged by our protocol requiring participating students to peer-assess the peer trainee following a structured framework which was meant to inform the subsequent feedback. Under these circumstances, they were unconsciously guided to provide
terminal feedback. Notwithstanding, and after reading the sample feedback narratives
given above, it would have been inappropriate to do so in front of the patients and as
instructed, students provided feedback privately. However, this terminal feedback is not
to be confused with the constructivist scaffolding provided during the procedure (Wood
et al., 1976; Vygotsky, 1978).

The second quantitative peer-feedback narrative analysis coded them according to their
sign as positive, negative or mixed. Once again, BDS 2 and BDS 5 students showed
notable differences. Thus, whilst the vast majority of both pre-clinical and clinical peer-
feedback narratives were coded as positive (91% for BDS 2 and 65% for BDS 5), BDS
5 negative ones were more than six fold those from BDS 2 (13% versus 2%). Similarly,
mixed sign narratives from clinical students were three fold those from pre-clinical ones
(22% versus 7%). These differences might be attributed to the contrasting pre-clinical
and clinical contexts in which peer-feedback was provided. “Real-life” clinical training
at dental schools involve being responsible for a patient, and so are very intensive and
demanding and students usually feel stressed and tired (Kossioni et al., 2012; Ostapczuk
et al., 2012; Tomás et al., 2013).

However, the higher proportions of negative and mixed narratives signs from clinical
students should not instinctively be considered damaging or unhelpful, especially if
their attention was focused to the “task” being assessed and not as a threat to the self-
esteeem (Kluger and DeNisi, 1996). Then, a possible explanation for a positive effect of
a negative signed peer-feedback might be found in the “Regulatory Focus Theory”
(Higgins, 1997), which extends the hedonistic principle that humans are motivated to
approach pleasure and avoid pain. Thus, Tory Higgins distinguishes two systems of
self-regulation. On the one hand, a “promotion focus” concerned with accomplishments
and aspirations which is motivated by achievement of rewards (things we want to do),
and on the other, a “prevention focus” concerned with safety and responsibilities that is
driven by the avoidance of pain or punishment (things we have to do) (Higgins, 1997;
Watling et al., 2012b).

Subsequently, using this Regulatory Focus Theory, Kluger and Van Dijk (2010),
explained why “neither positive nor negative signed feedback affects performance in a
constant manner”. Consequently, they showed that a positive signed feedback
contributes to motivation and performance under a “promotion focus”, for example in
tasks requiring creativity. By contrast, a negative signed feedback is encouraging under
a “prevention focus”, for example in tasks requiring vigilance and attention to detail, but this is debilitating under “promotion focus”.

Furthermore, Van Dijk and Kluger (2011) reported a large effect size ($d=0.67$) in performance of positive feedback among people working on promotion tasks (e.g. initiating changes, planning a task, challenging decision making, creative problem solving, assimilating new technology, presenting various alternatives, and generating ideas), but they also described a hampering effect size ($d=-0.37$) in performance when a positive sign feedback was given to individuals working on prevention tasks (e.g. detecting errors, maintaining safety, quality control, supervising, and work scheduling).

Coincidently, the former effect size for positive feedback in promotion focus tasks ($d=0.67$) resembles the ones from those BDS 2 and BDS 5 students who completed 10 or more peer-assessment encounters got in their average of all examination ($d=0.80$ and $d=0.74$, respectively) described in Chapter 5 (Table 5-15 in page 180 for BDS 2 and Table 5-16 in page 181 for BDS 5). Moreover, the current Chapter’s BDS 2 facts in that 91% of their feedback narratives were coded as positive, 77% were task specific, and 81% with a clinical content, amalgamated together with Van Dijk and Kluger (2011) descriptions of promotion tasks, it is conceivable that these BDS 2 students worked mostly in a “promotion focus”. This because they were generally involved in tasks such as “initiating changes” from theory to practice, “planning a task” to successfully accomplish it during the pre-clinical session, and “assimilating new technology” like drilling handpieces, ultrasound and x-ray machines. If we further consider the results from Chapter 5 where the same BDS 2 students’ perceived their pre-clinical tasks complexity mostly as “moderate” (66.7%) and “high” (14.1%) (Table 5-2 in page 165), it is again imaginable that the meaning of the tasks being practiced would incline them to adopt a “promotion focus”.

Regarding BDS 5 students, the coding of their peer-feedback narratives (content, sign and specificity) as well as their perceived clinical tasks complexity suggests the adoption of mixed “promotion” and “prevention” foci, as mentioned by Kluger and Van Dijk (2010) for medical staff. Firstly, considering that the majority of their feedback narratives were coded as positive and task specific, diversified in content domain (Table 6-1), and that these students’ mostly perceived clinical tasks complexity as “moderate” (42.6%) and “high” (25.4%) (Chapter 5, Table 5-3 in page 166), it insinuates these senior students adopted a “promotion focus”. Furthermore, and according to Van Dijk and Kluger (2011) promotion tasks list, they were making “challenging decision” with
Professionalism, embracing a “problem solving” attitude for clinical struggles, “presenting various alternatives” while communicating with patients and trainers, “generating ideas” while writing a treatment plan, and practicing management when “planning tasks”.

Secondly, however, these BDS 5 students were also performing tasks “they have to do” adopting a “prevention focus” (Higgins, 1997). Accordingly, while treating their patients, and especially when performing those dental crowns, dentures and root canal treatments perceived as highly complex (Chapter 5, Table 5-3 in page 166), they must have focused on detecting clinical errors, maintaining their patients’ safety, managing the quality of the procedures and professionally supervising each other, described as prevention tasks (Van Dijk and Kluger, 2011). Expanding this further, a negative signed peer-feedback (as the 13% in BDS 5 students) for this tasks requiring vigilance and attention to details, might have been encouraging for the students working on a “prevention focus” as described by Kluger and Van Dijk (2010). Further, under this kind of circumstance, Watling et al. (2012b) suggested that negative feedback (criticising or correcting a behaviour) is not only acceptable but desirable and more valued than a positive one (affirming or reinforcing a behaviour).

The above discussed assumptions of the present study results are consistent with Kluger and Van Dijk (2010) judgment in that “medical staff are faced with a mix of prevention and promotion foci”, especially in the case of our clinical BDS 5 students. However, our evidence is in complete disagreement with the same researchers speculation that the medical “community seems to emphasise error avoidance, risk management and the minimising of losses over creativity and other promotion-focused goals”. Consequently, it is not credible to base these statements just on a PubMed search comparing references with “diagnostic error” and “error reduction” (90,000) to those with ‘creativity’ and ‘opportunity’ (60,000) (Kluger and Van Dijk, 2010). In contrast, our data is in agreement with the detailed qualitative study by Watling et al. (2012b) in which they found that 22 medical learners were not “inclined to emphasise one focus over the other”. Further, the same task often included both foci and played a significant role in the workplace.

The qualitative thematic analysis of BDS 5 students’ peer-feedback narratives was, as expected, more clinically focused and patient-centred than those from BDS 2 students. Both groups of students provided peer-feedback to each other in a wide range of
Clinical, Communication, Management and Professionalism themes relevant to their particular pre-clinical and clinical working environments. However, four themes emerged at both groups’ narrative analysis. These were efficiency, infection control, time management and working speed, signalling these to be ongoing learning skills in spite of the difference between both pre-clinical and clinical courses.

In a mixed qualitative - quantitative methods study by Nofziger et al. (2010), second and fourth year medical students provided narratives about their prior peer-feedback experiences where both groups expressed the same themes. However, second-year students reported higher interpersonal style feedback than fourth-year ones (71% versus 63%), whilst conflicting feedback was reported higher in senior students (7% versus 1%). Among similar reported themes from our subjects, these medical students remembered having received feedback on “being too quiet”, “harder challenges”, “overconfidence”, “knowing your limits”, “low self-esteem” and “low self-confidence”. Another study on third-year medical students this time by White and Sharma (2012), reported similar written peer-feedback themes as our clinical BDS 5 students. Among these, they mentioned information management, initiative & self-direction, asking questions to learn, confidence, compassion, and respect for others.

The third part of the present Chapter’s results where focus groups were conducted, attempted to clarify why only 26% of BDS 2 and 44% of BDS 5 students decided to take part in the peer-assessment and peer-feedback exercise, but at the same time provide some insight into why, for future improvements, the majority of BDS 2 (74%) and BDS 5 (56%) students decide not to do so.

Thus, among the first group of students - those who peer-assessed each other – some expressed their desire to improve and develop their clinical performance practicing the peer-assessment protocol. As described by Lurie et al. (2006), this could be a sign of senior students’ mature judgements, as well as a positive ongoing process of students’ developing their professionalism, described by Roff et al. (2011) as “proto-professionalism”.

Another theme expressed by students as a reason to participate in the protocol that deserves some discussion is “the need for more feedback”. The motives for this might be found in the results of the 2013 UK NSS. Thus and despite that a high 80% of the KCLDI “BDS Dentistry” and 88% of “BDS Dentistry Graduate/Professional Entry
Programme” surveyed students (UK average 70%), agreed with the statement “Feedback on my work has been prompt”, a lower 61% and 63%, respectively (UK average 64%), agreed that “Feedback on my work has helped me clarify things I did not understand” (Unistats, 2014). This could be interpreted as whilst students are being provided with prompt staff feedback it is not being perceived by some students as helpful. This has been described in the literature as an effect of feedback provided to passive recipients (Teunissen et al., 2009).

Hence, the option of participating in the peer-assessment protocol where peer-feedback was conceived more as an accessible communication, interaction or dialogue between peers (Nicol and Macfarlane-Dick, 2006; Carless, 2007; Nicol, 2010), could have been seen as an opportunity to fill this gap.

Peers’ trust was another frequently mentioned theme which helps to understand the high proportion of participating students that found the protocol acceptable and fair (85% of BDS 2 and BDS 5) as described in the previous Chapter (Table 5-10 in page 170). Similarly, Mann et al. (2011) described peer-feedback that took place in a relationship of mutual trust and respect, was definitely enhanced. Further, this positive atmosphere also contributes to explain the positive influence peer-feedback had in participating students’ reflective skills and examination performance. Thus, as concluded by Watling et al. (2012a), credibility is a key factor for feedback to be influential. As they stated, only feedback that survives a critical assessment of the source credibility is likely to impact students’ learning. If not credible, feedback will probably fall on deaf ears (Bing-You et al., 1997).

Students enjoyed teaching each other. This is probably related to the increased students’ motivation as a product of the cooperative learning approach which has been described as ideal for courses that require students to learn skills that require manual dexterity, knowledge and clinical reasoning (Asghar, 2009). Further, a positive peer-assessment and peer-feedback experience, despite its challenges (Walker, 2014), might have improved students’ belief in themselves, increasing their confidence and self-efficacy (Bandura, 1997). Moreover, students’ cooperation during the peer-assessment protocol might have benefitted the feedback receiver stimulating her or his reflection as discussed in Chapter 5, but also the provider as a peer-teacher. Hence, according to Dales’s Learning Pyramid (Dale, 1946), listening leads to 5% recall, whereas teaching others, leads to 80% recall (ten Cate and Durning, 2007).
Additionally, this enjoyable cooperative experience based on peer-assessment and peer-feedback that was immediate, constructive, dialogic and focused to improve the performed task, might be a good example of socially constructed learning reflecting Vygotskian schools of thought (Vygotsky, 1978; Topping, 2005). This interaction between peers in an amenable way (Asghar, 2009) would be responsible for the benefits on the cognitive development produced in both assessor and assessed students, over learning in isolation (Vygotsky, 1978).

Despite the above mentioned intrinsic motivation themes, there also were some students that honestly stated some extrinsic reasons to participate in the peer-assessment study as they were motivated by the Kindle Fire HD prize draw, the Research Certificate for their portfolios, and even some others that wanted to help the PhD student conducting the study. Besides helping to augment the study sample, these extrinsic reasons together with the above intrinsic ones, allow us to accept the second hypothesis of this Chapter: Participating students decided to take part in the peer-assessment protocol as a result of intrinsic and extrinsic motivations.

Those students who chose not to peer-assess each other – but attended the focus groups – presented a few reasons, mostly intrinsic but extrinsic as well, for not taking part in the peer-assessment protocol. This, besides allowing us to accept the third hypothesis of the Chapter in that “students who decided not to take part in the peer-assessment protocol did so as a result of intrinsic and extrinsic motivations”, gave us a helpful insight for future implementations of the protocol.

Consequently, non-participating students talked about their uncomfortable earlier experiences with feedback while following their previous university degrees. This was a surprising revelation but was concurrent with the large number of students who already held a university degree and were now following the BDS programme. Thus, a total 26 BDS 2 (17% of the class) students and 37 BDS 5 (24% of the class) were part of the Dentistry Graduate/Professional Entry Programme or Dentistry Entry Programme for Medical Graduates. In agreement with this, a qualitative study conducted by Beaumont et al. (2011), investigated the impact of prior experiences of peer-assessment in higher education, and found a great proportion of students with negative experiences relating to trust, competency and plagiarism. Further, in the clinical environment Quantrill and Tun (2012) reported more than half of the 41 surveyed foundation medical trainees were not satisfied with the quality of the received feedback when assessed by staff with three
different forms of workplace-based assessment. In line with this and even more dramatic, Sabey and Harris (2011) surveyed results of 52 specialist medical trainees using workplace-based assessment showed that though they valued feedback, this was lost as it often came too late, it was of poor quality, and doubtfully honest, particularly when given face to face. Further, just now Urquhart et al. (2014) reported fifth year undergraduate medical students’ feedback experiences as a process that happened ‘to’ them rather than ‘with’ them, one-way process, and metaphorically described “feedback as war”. In light of this, the reason stated for choosing not to participate in the peer-assessment protocol based on bad previous experiences appears reasonable.

Another theme identified by those students who did not participate in the peer-assessment protocol, related to their reluctance to take the responsibility of assessing their peers. This feeling was also described by Brown and Glasner (1999), who reported the “common belief that assessment is the teacher’s responsibility and some students resent being required to do the teacher’s dirty work”. Others felt unable to assess their peers which could be related to the Biggs and Tang (2011) statement in that peer-assessment can be stressful to some students. Both these themes probably denote a failure in the protocol’s explanation and training to the students, as the “learning” of the peer-assessment process should have been emphasised over the “assessment”.

However, this specific issue must not make us lose the overall view of the problems confronted by the current approach to feedback in higher education. Our results are further evidence supporting a need for a change in the quality (Sargeant et al., 2008; Pelgrim et al., 2012b), quantity (Nicol and Macfarlane-Dick, 2006; Quantrill and Tun, 2012; Urquhart et al., 2014) and general approach (Walsh et al., 2009; Archer, 2010; Hatala et al., 2014) to improve the provision of a much needed feedback. In this perspective, this Chapter results have shown that peers can be of great assistance in providing valuable feedback to change students’ behaviour.

6.7 Conclusion

This Chapter’s analysis of peer-feedback narratives has shown that both pre-clinical as well as clinical dental undergraduate students, using their respective peer-DOPS and peer-mCEX forms as frameworks, were able to engage in a dialogic feedback focused to different domains, and of contrasting signs and specificities. However, while pre-clinical students’ feedbacks were mostly focused on the Clinical domain, positive and
task specific, clinical ones were more concerned with Management & Leadership and Communication skills, and at the same time dared to provide more negative or mixed feedback keeping it task specific.

Thus, as emerged narrative themes resembled those differentiated items from their respective pre-clinical and clinical peer-assessment forms, it appears these provided a framework to help students in the provision of different feedback codes according to their specific work based circumstances.

Training on peer-assessment and peer-feedback should emphasise a collaborative and formative approach to learning in order to integrate those students with previous negative experiences with feedback. Further, the protocol should include intrinsic as well as extrinsic motivations in order to increase students’ participation.
General Discussion

The educational literature supports peer-assessment and particularly its feedback component as a way of fostering student learning. The development, piloting and implementation of the current peer-assessment protocol demonstrated positive results which are summarised below (Table GD-1). However, despite its feasibility, peer-assessment is generally lacking in undergraduate education (Boud, 1991).

Based on Barr’s adaptation (2000) of the four levels of educational intervention’s hierarchy (Levels 1, 2a, 2b, 3, 4a and 4b) according to Kirkpatrick (1967), the academic effects of the studied peer-assessment protocol in students who completed 10 or more peer encounters can be coded as Level 2b: Acquisition of knowledge/skills. The hierarchy relates this to the “…acquisition of concepts, procedures and principles of inter-professional collaboration”, and/or the “…attainment of thinking/problem-solving, psychomotor and social skills linked to collaboration” (Barr et al., 2000 p. 10).

Similarly, the development of students’ reflective skills allows this part of the intervention to be categorised as Level 3: Change in behaviour, which implies a “behavioural change transferred from the learning environment to the workplace prompted by modifications in attitudes or perceptions, or the application of newly acquired knowledge/skills in practice” (Barr et al., 2000 p. 11).

These findings are important as most studies in medicine and dentistry provide educational outcomes limited to Level 1 (learner’s reactions) and Level 2 (2a - modification of attitudes and perceptions, and 2b -acquisition of knowledge/skills), and only occasionally Level 3 (Change in behaviour) (Gordon and Findley, 2011; Hendricson, 2012). In addition, a review on learner outcomes in healthcare interprofessional continuing education (Gillan et al., 2011) found 22% of the outcomes at Level 1; 66% at Level 2a; 1.4% at Level 2b; 10.2% at Level 3; and only 0.4% at Level 4 (4a -change in organisational practice, and 4b -benefits to patients).

Pedagogically, the most important contribution of the present thesis is the evidence that a carefully and rigorously planned longitudinal peer-assessment and peer-feedback protocol of dental undergraduate students can significantly enhance their academic achievements and develop their reflective skills (Chapter 5). Using Biggs and Tang adagio (2011 p. 51), the former allows them to fish for today’s meal; the latter is the net that provides meals for the rest of their lives.
Table GD-1 Aims and outcome summary for each research chapter of the thesis.

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Aims</th>
<th>Outcome summary</th>
<th>Main implications</th>
</tr>
</thead>
</table>
| 2       | • To investigate the different teaching and assessment methodologies used in dental education using KCLDI as a true model. | • KCLDI comprises 1716 teaching sessions with the majority being lectures (40%) and practical & clinical (38%).  
• Assessment includes 23 different methodologies and 153 sessions (55% formative, 30% summative, 15% hurdle).  
• OSCE was the most commonly used assessment methodology (53% of institutions), followed by MCQ (40%), preclinical observation (32%), and WPBA (32%).  
• Peer-assessment was used by 11% of the institutions. | • Similar assessment strategies are shared by dental teaching institutions from different countries and continents.  
• New clinical assessment techniques are being employed in undergraduate dental education.  
• The proportion of methods that assess critical thinking and problem solving is small. |
| 3       | • To develop and pilot a structured protocol of formative, prospective peer-assessment as an informed framework for immediate peer-feedback. | • The developed peer-assessment protocol is reliable, feasible, and well accepted by students. | • It is worth implementing the protocol in a larger sample followed for a longer period to provide generalizable data. |
| 4       | • To assess the reliability and construct validity of a self-reported reflection questionnaire as a method of assessing dental students’ and postgraduate trainees’ reflection habits. | • Kember’s quantitative Reflection Questionnaire is a valid and reliable instrument to assess dental students’ reflective skills.  
• The higher the academic ladder the more reflective students are.  
• Those ≥ 24 years of age and with a previous university degree demonstrated higher reflective habits.  
• There is no gender difference in reflective skills. | • The diagnostic reflection tool can be used as pre- and post-intervention of the peer-assessment and peer-feedback protocol. |
| 5       | • To implement and assess the reliability, validity, feasibility, acceptability and educational impact of the peer-assessment and peer-feedback protocol. | • The protocol is reliable, valid, feasible and acceptable.  
• It has the potential to make a significant educational impact, provided it is practiced during ten or more peer encounters. | • Dental students’ academic and reflective skills can be developed through a structured and longitudinal peer-assessment and peer-feedback protocol, provided it is exercised at least ten times during the academic year. |
| 6       | • To quantitatively and qualitatively analyse students’ written peer-feedback narratives. | • Pre-clinical students’ feedbacks were mostly Clinical, positive and task specific.  
• Clinical ones were more concerned with Management & Leadership and Communication skills, and provided more negative or mixed feedback keeping it task specific.  
• Bad previous peer-assessment experiences precluded students to participate in the peer-assessment protocol. | • Peer-assessment framework help students to provide feedback focused to different domains, and of contrasting signs and specificities.  
• Students’ training on peer-assessment and peer-feedback must stress its collaborative and formative approach to integrate students with previous negative experiences.  
• Intrinsic and extrinsic motivations to participate should be included.  
• To scrutinise the reasons why students decided to take part in, or not take part in the peer-assessment protocol. |
The enhancement of students’ skills is the result of an intervention that deliberately intended to increase the scope, frequency and, if possible, the depth of participants’ feedback, to complement feedback tutors can reasonably provide (Evans, 2013). The powerful effects of such peer-feedback protocol encourage the organisation of collaborative student working (Ladyshewsky, 2013).

Taken as a whole, the results of the peer-assessment and peer-feedback protocol meets the criteria that effective learning needs to be engaging, contextualised, active, and social (Chickering and Gamson, 1999; Forsyth, 2003; Kaufman, 2003; Biggs and Tang, 2011; Harden and Laidlaw, 2013).

The peer-assessment protocol achieved this by requiring students to engage in two different ways:

In a constructivist model they connected and engaged the actual concrete experience of simulation (pre-clinical BDS 2 students) and clinical performance (clinical BDS 5 students) to previous similar events to activate what they already knew, had experienced and felt (Dennick, 2012; Pritchard, 2013).

It resulted in “observing students” engaging with the task of their peers. For example, the observing students classified the complexity of each procedure (Chapter 5, Table 5-2 in page 165 and Table 5-3 in page 166), accurately marked their peers performance (Chapter 5, Table 5-11 in page 171), and provided a “task specific” feedback in 77% of BDS 2 and 74% of BDS 5 encounters (Chapter 6, Table 6-1 in page 205). For them to set the case complexity, reliably assess their peers, and provide detailed feedback narratives they had to engage in their peers’ tasks. It is also arguable that all students - participating and non-participating - were required to engage in their own respective tasks.

A possible explanation for this positive engagement effect might be the alignment (Biggs and Tang, 2011) of the peer-assessment domains, from both peer-DOPS and peer-mCEX, with the students’ respective course learning outcomes and therefore their formal assessment (Nofziger et al. (2010), as shown in Chapter 3 (Table 3-1, pages 115-117). In contrast, a lack of alignment might have led to not engaging with the peer-feedback protocol (Ladyshewsky, 2010).

Similarly, the need for learning to be contextualised was addressed as both groups of participating students worked situated, be it in a pre-clinical simulation or in clinical
practice, in what are the day-to-day dental procedures. Whilst non-participating students were also doing this, peer-feedback narrative analysis from Chapter 6 demonstrates how the assessment domains of each of the BDS 2 and BDS 5 forms described in Chapter 3, might have helped both observing and performing students to contextualise the task itself not only as the required pre-clinical or clinical procedure but as a broader learning opportunity.

Consequently, it is hard to see if it was not for the peer-DOPS or peer-mCEX frameworks, how BDS 2 students put the very pre-clinical task in context, for example with working speed, time management and communication (Chapter 6, Table 6-2 in page 207). Further, BDS 5 students spoke about knowing their limits, efficiency, professional attitude and anticipating events (Chapter 6, Table 6-3 in page 208). This supports the idea that the peer-assessment experience demanded the students to be active contributors to their learning process as none in the working pairs could be simple passive spectators (Chickering and Gamson, 1999). It is well known that active learning results in a deeper processing of the learning experience (Divaris et al., 2008; Harden and Laidlaw, 2013; Nicol et al., 2014).

Moreover, by actively observing, scaffolding, assessing and providing feedback to each other in a social constructivist model (stressing social interaction), they had to ‘share’ what they know, what they don’t know, relate it to previous pre-clinical or clinical experiences and apply it to the current simulation or real patient treatment. As a result, they learnt by awaking developmental processes that only operate when interacting with their peers (Vygotsky, 1978). Further, the feedback was timely and immediate which, as discussed in Chapter 1 (page 71), is an important factor for effective feedback (Subramanian et al., 2013), particularly for novices who need to learn things quickly, in context, and when still engaged with the task (Ladyshewsky, 2013). Students who were not involved in the study similarly had to actively work in their tasks but the difference was the relative absence of ‘share talk’ (Vygotsky, 1978) or dialogic feedback (Beaumont et al., 2011; Pritchard, 2013) participating pre-clinical and clinical partners performed during and after their tasks.

In addition, whilst all students (both BDS 2 and BDS 5) were “actively engaged” in their own pre-clinical or clinical tasks, those who participated in the peer-assessment exercise benefited from the effects of being both observer (assessor) and trainee. As discussed in Chapter 1, previous research has demonstrated the benefits of peer-
feedback for the peer-trainee (Falchikov and Goldfinch, 2000). However, the active social engagement in the pre-clinical or clinical tasks also produces cognitive gains for the peer-assessor (Topping, 1996). This occurs as they require an extra effort to ‘understand’ what has happened during the task to be able to feed back to the peer-trainee in simple terms (Ladyshewsky, 2013). In other words, there is a cognitive demand on the peer-assessor to monitor, detect, diagnose and correct her/his peer’s performance that produces the learning benefits (Topping, 2005).

Similarly, the effect on peer-trainees are based on the self-reflection that comes from the non-evaluative dialogic peer-feedback (Ladyshewsky, 2013). The social interaction can therefore transform pre-clinical or clinical experiences into reflective learning which encourages reflective practice (Matheson, 2008). This is an important attribute for dental practitioners who deal with complex professional contexts such as exponential new knowledge and treatment possibilities, more stringent and informed multicultural patients in an increased multidisciplinary focus of practice (Koole et al., 2013). Reflection requires dedicated time (Jonas-Dwyer et al., 2013) which may in part explain its underutilisation as a teaching strategy (Boyd, 2002).
Conclusion

The development of a peer-assessment and peer-feedback protocol has demonstrated its short-term potential to boost students’ academic learning and achievement outcomes as well as their reflective skills. However, we would need to follow-up the current students to evaluate the benefits to their professional careers in the medium and long-term.
Future Studies

It is suggested that future studies might involve:

- Evaluation of the effects of the current peer-assessment protocol in other dental student levels such as BDS 3 and BDS 4.

- Assessment of the outcomes of the current protocol implemented in “group practice teams” where students of different BDS levels work together in the treatment of patients. Thus, near peers would assess each other according to their levels of experience.

- Evaluation of the effect of organising “group practice teams” or pairs of students with different levels of reflection skills, thinking that low reflective ones might benefit from high reflective peers.

- Appraisal of the results of implementing the peer-assessment protocol in a multidisciplinary environment where students of different professions work collaboratively together in order to provide a better service to patients. An ideal setting for this might be the outreach “Integrated Team Care Education” programme KCLDI students complete at the University of Portsmouth Dental Academy.

- Replication of the studies in other national or international health-related teaching institutions.

- Repeating the studies in the postgraduate arena. Both within dentistry and other health-related professions.
References


Ambrose, L. & Ker, J. 2014. Levels of Reflective Thinking and Patient Safety: An


References

Teacher, 35, e984-e989.


References


References


References


Crossley, J., Johnson, G., Booth, J. & Wade, W. 2011. Good Questions, Good Answers:
Construct Alignment Improves the Performance of Workplace-Based Assessment Scales. *Medical Education*, 45, 560-569.


Denehy, G. & Fuller, J. 1974. Student Peer Evaluation: An Adjunct to Preclinical


References

Human Intuition and Expertise in the Era of the Computer, New York, USA, Free Press.
Improvement Quarterly, 6, 50-72.


and Tutor Feedback. *Assessment & Evaluation in Higher Education*, Early Online.


Hatton, N. & Smith, D. 1995. Reflection in Teacher Education: Towards Definition and
References


and Their Influence on Teaching and Learning. *Medical Education*, 38, 448-454.


References

a Comparative Study between Faculty Members and Pre-Clinical Students. *European Journal of Dental Education*, 15, 250-254.


King's College London Dental Institute. 2012. *Bds Programme Description* [Online]. KCLDI.


Knowles, M., Holton, E. & Swanson, R. 2012. The Adult Learner: The Definitive


Lurie, S., Nofziger, A., Meldrum, S., Mooney, C. & Epstein, R. 2006. Temporal and


References


References


Oliver, R., Kersten, H., Vinkka-Puhakka, H., Alpasan, G., Bearn, D., Cema, I., Delap,
References


Planas Lladó, A., Soley, L., Fraguell Sansbelló, R., Pujolras, G., Planella, J., Roura-


References


York, USA, Basic Books.


Tolsgaard, M., Arendrup, H., Pedersen, P. & Ringsted, C. 2013. Feasibility of Self-


Watson, J. 1913. Psychology as the Behaviorist Views It. *Psychological Review*, 20,
158-177.


Wingate, U. 2010. The Impact of Formative Feedback on the Development of


